

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C07H 17/00, C07K 2/00, 14/00, C12N 15/00, C12Q 1/68	A1	(11) International Publication Number: WO 95/16694 (43) International Publication Date: 22 June 1995 (22.06.95)
(21) International Application Number: PCT/US94/14563 (22) International Filing Date: 16 December 1994 (16.12.94) (30) Priority Data: 08/168,479 16 December 1993 (16.12.93) US (71) Applicants: COLD SPRING HARBOR LABORATORY [US/US]; One Bungtown Road, Cold Spring Harbor, NY 11724 (US). THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 22nd floor, 300 Lakeside Drive, Oakland, CA 94612-3550 (US).		(72) Inventors: STILLMAN, Bruce, W.; 21 Carnegie Avenue, Cold Spring Harbor, NY 11724 (US). BELL, Stephen, P.; Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724-0100 (US). KOBAYASHI, Ryuji; Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724-0100 (US). RINE, Jasper; 2948 Pine Avenue, Berkeley, CA 94704 (US). FOSS, Margit; Duke University Medical Center, Dept. of Cell Biology, Durham, NC 27710 (US). McNALLY, Francis, J.; 365 Valdez Avenue, San Francisco, CA 94127 (US). LAURENSEN, Patricia; 1235 Tenth Avenue, San Francisco, CA 94122 (US). HERSKOWITZ, Ira; University of California, Dept. of Biochemistry and Biophysics, San Francisco, CA 94143 (US). LI, Joachim, J.; 700 Parnassus Avenue #18, San Francisco, CA 94122 (US). (74) Agents: OSMAN, Richard. Aron et al.; Flehr Hobbach Test Albritton & Herbert, Suite 3400, 4 Embarcadero Center, San Francisco, CA 94111-4187 (US). (81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: ORIGIN OF REPLICATION COMPLEX GENES, PROTEINS AND METHODS (57) Abstract Origin of DNA Replication Complex (ORC) genes, recombinant ORC peptides and methods of identifying DNA binding proteins and using the subject compositions are provided. Vectors and cells comprising such ORC genes find use in the production of recombinant ORC peptides. The subject ORC peptides find particular use in screening for ORC selective agents useful in the diagnosis, prognosis or treatment of disease, particularly fungal infections and neoproliferative disease. Disclosed methods for identifying a gene encoding a protein which directly or indirectly associates with a selected DNA sequence involve: transforming an expression library of hybrid proteins into a reporter strain, wherein the library comprises protein-coding sequences fused to a constitutively expressed transcription activation domain and the reporter strain comprises a reporter gene with at least one copy of a selected DNA sequence in its promoter region. Clones expressing the transcription or translation product of the reporter gene are detected and recovered.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Larvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

*ORIGIN OF REPLICATION COMPLEX GENES,
PROTEINS AND METHODS*

INTRODUCTION

The research carried out in the subject application was supported in part by grants from the National Institutes of Health. The government may have rights in any patent issuing on this application.

5

Technical Field

The technical field of this invention concerns Origin of Replication Complex genes which are involved with DNA transcription and replication.

10

Background

The elements involved in the early events of eukaryotic DNA replication have begun to emerge in the yeast *Saccharomyces cerevisiae*. A critical first step was the identification of ARS elements derived from yeast chromosomes, a subset of which were subsequently shown to act as chromosomal origins of DNA replication (reviewed in 11). Sequence comparison of a number of ARS elements resulted in the identification of the ARS consensus sequence (ACS, 12). This sequence is essential for the function of yeast origins of DNA replication (7, 12, 13). Three additional elements required for efficient ARS1 function have been identified. When mutated individually, these elements, referred to as B1, B2, and B3, result in a slight reduction of ARS1 activity. When two or three of the B elements are simultaneously mutated, however, ARS1 function is severely compromised (14).

15
20

Proteins that recognize two elements of ARS1 have been identified. The yeast transcription factor ABF1 binds to and mediates the function of the B3 element (11, 14). More recently we have identified a multi-protein complex that specifically recognizes the highly conserved ACS (15). This activity, referred to as the origin recognition complex (ORC), has several properties that make it an attractive candidate to act as an initiator protein at yeast origins of replication. Binding of this protein requires the ACS, and the effect of mutations in the consensus sequence on ARS1 function parallels the effect of the same mutations on ORC DNA binding. ORC binds to more than 10 yeast ARS elements, several of which are known origins of DNA replication (15). Specific DNA binding by ORC requires ATP, suggesting that ORC binds ATP, a property of a number of known initiator proteins (17). ORC also interacts with other sequences outside of the ACS that are known to be important for ARS function (18, 19). Further support for the hypothesis that ORC mediates the function of the ACS is provided by in situ deoxyribonuclease I (DNase I) footprinting experiments that identify a protected region of ARS1 remarkably similar to that observed with ORC in vitro (20).

Relevant Literature

A multi-protein complex that recognizes cellular origins of DNA replication was reported in Bell and Stillman (1992) Nature 357, 128-134. Much of the present disclosure was published by Foss et al. (1993), Bell et al. (1993) and Li and Herskowitz (1993), in Science 262, 1838, 1843 and 1870, respectively, issue date December 17, 1993. Wang and Reed (1993) Nature 364, 121-126 report using a single-hybrid screen as disclosed herein.

SUMMARY OF THE INVENTION

Origin of DNA Replication Complex (ORC) genes, recombinant ORC peptides and methods of identifying DNA binding proteins and using the subject compositions are provided.

Provided are compositions comprising isolated nucleic acids encoding unique ORC gene portions, especially portions encoding biologically active unique portions of ORC1-ORC6 proteins. Vectors and cells comprising such DNA molecules find use in the production of recombinant ORC peptides.

The subject compositions are used to isolate ORC genes from a wide variety of species, including human. The subject ORC peptides also find particular use in screening for ORC selective agents useful in the diagnosis, prognosis or treatment of disease, particularly fungal infections and neoproliferative disease.

- 5 Particularly useful are agents capable of distinguishing an ORC protein of an infectious organism or transformed cell from the wild-type human homologue.

Also disclosed are methods for identifying a gene encoding a protein which directly or indirectly associates with a selected DNA sequence. Generally, the methods involve transforming an expression library of hybrid proteins into a
10 reporter strain, wherein the library comprises protein-coding sequences fused to a constitutively expressed transcription activation domain and the reporter strain comprises a reporter gene with at least one copy of a selected DNA sequence in its promoter region. Clones expressing the transcription or translation product of the reporter gene are detected and recovered. A preferred method employs an
15 activation domain from GAL4 and a lacZ reporter gene.

BREIF DESCRIPTION OF SEQUENCE ID NUMBERS

SEQUENCE ID NO:1. DNA Sequence of *ORC1*.

SEQUENCE ID NO:2. Amino Acid Sequence of *ORC1*.

- 20 SEQUENCE ID NO:3. DNA Sequence of *ORC2*.

SEQUENCE ID NO:4. Amino Acid Sequence of *ORC2*.

SEQUENCE ID NO:5. DNA Sequence of *ORC3*.

SEQUENCE ID NO:6. Amino Acid Sequence of *ORC3*.

SEQUENCE ID NO:7. DNA Sequence of *ORC4*.

- 25 SEQUENCE ID NO:8. Amino Acid Sequence of *ORC4*.

SEQUENCE ID NO:9. DNA Sequence of *ORC5*.

SEQUENCE ID NO:10. Amino Acid Sequence of *ORC5*.

SEQUENCE ID NO:11. DNA Sequence of *ORC6*.

SEQUENCE ID NO:12. Amino Acid Sequence of *ORC6*.

30

DESCRIPTION OF SPECIFIC EMBODIMENTS

The recombinant polypeptides of the invention comprise unique portions of the disclosed ORC proteins which retain an binding affinity specific to the subject

full-length ORC protein. A "unique portion" has an amino acid sequence unique to subject ORC in that it is not found in previously known protein and has a length at least long enough to define a peptide specific to that ORC. Unique portions are found to vary from about 5 to about 25 residues, usually from 5 to 10 residues in length, depending on the particular amino acid sequence and are readily identified by comparing the subject portion sequences with known peptide/protein sequence data bases. Hence, the term polypeptide as used herein defines an amino acid polymer with as few as five residues. ORCs used in the subject screening assays are frequently smaller deletion mutants of full-length ORC proteins. Typically, such deletion mutants are readily generated using conventional molecular techniques and screened for an ORC-specific binding affinity using the various assays described below, e.g. footprint analysis, coimmunoprecipitation, etc.

ORC-specific retained binding affinities include the ability to selectively bind a nucleic acid of a defined sequence, an ORC protein or an compound such as an antibody which is capable of selectively binding an ORC protein. As such, binding specificity may be provided by an ORC-specific immunological epitope, lectin binding site, etc. Selective binding is conveniently shown by competition with labeled ligand using recombinant ORC peptide either in vitro or in cell based systems as disclosed herein. Generally, selective binding requires a binding affinity of 10^{-6} M, preferably 10^{-8} M, more preferably 10^{-10} M, under in vitro conditions as exemplified below.

The subject recombinant polypeptides may be free or covalently coupled to other atoms or molecules. Frequently the polypeptides are present as a portion of a larger polypeptide comprising the subject polypeptide where the remainder of the larger polypeptide need not be ORC-derived. The subject polypeptides are typically "isolated", meaning unaccompanied by at least some of the material with which they are associated in their natural state. Generally, an isolated polypeptide constitutes at least about 1%, preferably at least about 10%, and more preferably at least about 50% by weight of the total poly/peptide in a given sample. By pure peptidepolypeptide is intended at least about 60%, preferably at least 80%, and more preferably at least about 90% by weight of total polypeptide. Included in the subject polypeptide weight are any atoms, molecules, groups, etc. covalently

coupled to the subject polypeptides, such as detectable labels, glycosylations, phosphorylations, etc.

The subject polypeptides may be isolated or purified in a variety of ways known to those skilled in the art depending on what other components are present in the sample and to what, if anything, the polypeptide is covalently linked. Purification methods include electrophoretic, molecular, immunological and chromatographic techniques, especially affinity chromatography and RP-HPLC in the case of peptides. For general guidance in suitable purification techniques, see Scopes, R., Protein Purification, Springer-Verlag, NY (1982).

The polypeptides may be modified or joined to other compounds using physical, chemical, and molecular techniques disclosed or cited herein or otherwise known to those skilled in the relevant art to affect their ORC/receptor binding specificity or other properties such as solubility, membrane transportability, stability, toxicity, bioavailability, localization, detectability, in vivo half-life, etc. as assayed by methods disclosed herein or otherwise known to those of ordinary skill in the art. Other modifications to further modulate binding specificity/affinity include chemical/enzymatic intervention (e.g. fatty acid-acylation, proteolysis, glycosylation) and especially where the poly/peptide is integrated into a larger polypeptide, selection of a particular expression host, etc. Amino and/or carboxyl termini may be functionalized e.g., for the amino group, acylation or alkylation, and for the carboxyl group, esterification or amidification, or the like.

Many of the disclosed poly/peptides contain glycosylation sites and patterns which may be disrupted or modified, e.g. by enzymes like glycosidases. For instance, N or O-linked glycosylation sites of the disclosed poly/peptides may be deleted or substituted for by another basic amino acid such as Lys or His for N-linked glycosylation alterations, or deletions or polar substitutions are introduced at Ser and Thr residues for modulating O-linked glycosylation. Glycosylation variants are also produced by selecting appropriate host cells, e.g. yeast, insect, or various mammalian cells, or by in vitro methods such as neuraminidase digestion. Other covalent modifications of the disclosed poly/peptides may be introduced by reacting the targeted amino acid residues with an organic derivatizing (e.g. methyl-3-[(p-azido-phenyl)dithio] propioimide) or crosslinking agent (e.g. 1,1-bis(diazoacetyl)-2-phenylethane) capable of reacting with selected side chains or

termini. For therapeutic and diagnostic localization, the subject poly/peptides thereof may be labeled directly (radioisotopes, fluorescers, etc.) or indirectly with an agent capable of providing a detectable signal, for example, a heart muscle kinase labeling site.

- 5 ORC polypeptides with ORC binding specificity are identified by a variety of ways including crosslinking, or preferably, by screening such polypeptides for binding to or disruption of ORC-ORC complexes. Additional ORC-specific agents include specific antibodies that can be modified to a monovalent form, such as Fab, Fab', or Fv, specifically binding oligopeptides or oligonucleotides and most
10 preferably, small molecular weight organic compounds. For example, the disclosed ORC peptides are used as immunogens to generate specific polyclonal or monoclonal antibodies. See, Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Laboratory, for general methods.

- Other prospective ORC specific agents are screened from large libraries of
15 synthetic or natural compounds. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily producible. Additionally, natural and synthetically produced libraries and compounds are readily modified through conventional chemical, physical, and biochemical means. See, e.g. Houghten et al. and Lam et al (1991) *Nature* 354,
20 84 and 81, respectively and Blake and Litzi-Davis (1992), *Bioconjugate Chem* 3, 510.

- Useful agents are identified with assays employing a compound comprising the subject polypeptides or encoding nucleic acids. A wide variety of in vitro, cell-free binding assays, especially assays for specific binding to immobilized
25 compounds comprising ORC polypeptide find convenient use. For example, immobilized ORC-ORC or ORC-nucleic acid complexes provide convenient targets for disruption, e.g. as measured by the disassociation of a labelled component of the complex. Such assays are amenable to scale-up, high throughput usage suitable for volume drug screening. While less preferred, cell-based assays may be used to
30 determine specific effects of prospective agents.

 Preferred agents are ORC- and species-specific. Useful agents may be found within numerous chemical classes, though typically they are organic compounds; preferably small organic compounds. Small organic compounds have

a molecular weight of more than 150 yet less than about 4,500, preferably less than about 1500, more preferably, less than about 500. Exemplary classes include steroids, heterocyclics, polycyclics, substituted aromatic compounds, and the like.

Selected agents may be modified to enhance efficacy, stability, pharmaceutical compatibility, and the like. Structural identification of an agent may be used to identify, generate, or screen additional agents. For example, where peptide agents are identified, they may be modified in a variety of ways as described above, e.g. to enhance their proteolytic stability. Other methods of stabilization may include encapsulation, for example, in liposomes, etc. The subject binding agents are prepared in any convenient way known to those in the art.

For therapeutic uses, the compositions and agents disclosed herein may be administered by any convenient way. Small organics are preferably administered orally; other compositions and agents are preferably administered parenterally, conveniently in a pharmaceutically or physiologically acceptable carrier, e.g., phosphate buffered saline, or the like. Typically, the compositions are added to a retained physiological fluid. As examples, many of the disclosed therapeutics are amenable to direct injection or infusion, topical, intratracheal/nasal administration e.g. through aerosol, intraocularly, or within/on implants e.g. collagen, osmotic pumps, grafts comprising appropriately transformed cells, etc. Generally, the amount administered will be empirically determined, typically in the range of about 10 to 1000 $\mu\text{g/kg}$ of the recipient. For peptide agents, the concentration will generally be in the range of about 50 to 500 $\mu\text{g/ml}$ in the dose administered. Other additives may be included, such as stabilizers, bactericides, etc. These additives will be present in conventional amounts.

The invention provides isolated nucleic acids encoding ORC genes, their transcriptional regulatory regions and the disclosed unique ORC polypeptides which retain ORC-specific function. As used herein: an "isolated" nucleic acid is present as other than a naturally occurring chromosome or transcript in its natural state and is typically joined in sequence to at least one nucleotide with which it is not normally associated on a natural chromosome; nucleic acids with substantial sequence similarity hybridize under low stringency conditions, for example, at 50°C and SSC (0.9 M saline/0.09 M sodium citrate) and remain bound when

subject to washing at 55°C with SSC, wherein regions of non-identity of substantially similar nucleic acid sequences preferably encode redundant codons; a partially pure nucleotide sequence constitutes at least about 5%, preferably at least about 30%, and more preferably at least about 90% by weight of total nucleic acid present in a given fraction; unique portions of the disclosed nucleic acids are of length sufficient to distinguish previously known nucleic acids, hence a unique portion has a nucleotide sequence at least long enough to define a novel oligonucleotide, usually at least about 18 bp in length, preferably at least about 36 nucleotides in length.

Typically, the invention's ORC polypeptide encoding polynucleotides are associated with heterologous sequences. Examples of such heterologous sequences include regulatory sequences such as promoters, enhancers, response elements, signal sequences, polyadenylation sequences, etc., introns, 5' and 3' noncoding regions, etc. According to a particular embodiment of the invention, portions of the coding sequence are spliced with heterologous sequences to produce soluble, secreted fusion proteins, using appropriate signal sequences and optionally, a fusion partner such as β -Gal. For antisense applications where the inhibition of expression is indicated, especially useful oligonucleotides are between about 10 and 30 nucleotides in length and include sequences surrounding the disclosed ATG start site, especially the oligonucleotides defined by the disclosed sequence beginning about 5 nucleotides before the start site and ending about 10 nucleotides after the disclosed start site. The ORC encoding nucleic acids can be subject to alternative purification, synthesis, modification, sequencing, expression, transfection, administration or other use by methods disclosed in standard manuals such as Current Protocols in Molecular Biology (Eds. Ausubel, Brent, Kingston, More, Feidman, Smith and Stuhl, Greene Publ. Assoc., Wiley-Interscience, NY, NY, 1992) or that are otherwise known in the art.

The invention also provides vectors comprising the described ORC nucleic acids. A large number of vectors, including plasmid and viral vectors, have been described for expression in a variety of eukaryotic and prokaryotic hosts. Advantageously, vectors will often include a promotor operably linked to an ORC polypeptide-encoding portion, one or more replication systems for cloning or expression, one or more markers for selection in the host, e.g. antibiotic

resistance. The inserted coding sequences may be synthesized, isolated from natural sources, prepared as hybrids, etc. Suitable host cells may be transformed/transfected/infected by any suitable method including electroporation, CaCl_2 mediated DNA uptake, viral infection, microinjection, microprojectile, or
5 other methods.

Appropriate host cells include bacteria, archebacteria, fungi, especially yeast, and plant and animal cells, especially mammalian cells. Of particular interest are E. coli, B. subtilis, Saccharomyces cerevisiae, SF9 cells, C129 cells, 293 cells, Neurospora, and CHO, COS, HeLa cells, immortalized mammalian
10 myeloid and lymphoid cell lines, and pluripotent cells, especially mammalian ES cells and zygotes. Preferred expression systems include COS-7, 293, BHK, CHO, TM4, CV1, VERO-76, HELA, MDCK, BRL 3A, W138, Hep G2, MMT 060562, TRI cells, and baculovirus systems. Preferred replication systems include M13, ColE1, SV40, baculovirus, lambda, adenovirus, AAV, BPV, etc. A large number
15 of transcription initiation and termination regulatory regions have been isolated and shown to be effective in the transcription and translation of heterologous proteins in the various hosts. Examples of these regions, methods of isolation, manner of manipulation, etc. are known in the art.

For the production of stably transformed cells and transgenic animals, the
20 subject nucleic acids may be integrated into a host genome by recombination events. For example, such a nucleic acid can be electroporated into a cell, and thereby effect homologous recombination at the site of an endogenous gene, an analog or pseudogene thereof, or a sequence with substantial identity to an ORC-encoding gene. Other recombination-based methods such as nonhomologous
25 recombinations, deletion of endogenous gene by homologous recombination, especially in pluripotent cells, etc., provide additional applications. Preferred transgenics and stable transformants over-express or under-express (e.g. knock-out cells and animals) a disclosed ORC gene and find use in drug development and as a disease model. Methods for making transgenic animals, usually rodents, from
30 ES cells or zygotes are known to those skilled in the art.

The compositions and methods disclosed herein may be used to effect gene therapy. See, e.g. Zhu et al. (1993) Science 261, 209-211; Gutierrez et al. (1992) Lancet 339, 715-721. For example, cells are transfected with ORC-encoding

sequences operably linked to gene regulatory sequences capable of effecting altered ORC expression or regulation. To modulate ORC translation, target cells may be transfected with complementary antisense polynucleotides. For gene therapy involving the grafting/implanting/transfusion of transfected cells, administration
5 will depend on a number of variables that are ascertained empirically. For example, the number of cells will vary depending on the stability of the transferred cells. Transfer media is typically a buffered saline solution or other pharmacologically acceptable solution. Similarly the amount of other administered compositions, e.g. transfected nucleic acid, protein, etc., will depend on the
10 manner of administration, purpose of the therapy, and the like.

The genes encoding six ORC subunits from *S. cerevisiae* are used to obtain the functional homologues of the ORC proteins from other species. For example, we have demonstrated that the *ORC1* gene is conserved in a related fungi *klyuermyses lactis*. The *ORC1* gene in both *S. cerevisiae* and *k lactis* contain
15 conserved primary protein sequence that are utilized to obtain the *ORC1* gene from other species including other fungi and from human. Using oligonucleotide primers based on the conserved sequences between *S. cerevisiae* and *k lactis*, PCR is used to identify the *ORC1* protein in any eukaryotic species. The cloned gene encoding *ORC1* polypeptide from any fungi or from human cells is used to express
20 the protein in a bacterial expression system to make antibodies against the polypeptide. These antibodies are used to immunoprecipitate the ORC complex from the relevant species. Using the disclosed techniques for protein sequencing, the sequence the ORC polypeptides is obtained. Using the protein sequencing methodologies disclosed herein for cloning the *S. cerevisiae* protein, other genes or
25 cDNAs encoding the ORC polypeptides from other fungi species and from human cells are obtained. As we demonstrate herein how to reconstitute the ORC complex by expressing each of the *S. cerevisiae* genes in a baculovirus expression vector and infecting Sf 9 insect cells with viruses expressing each of the ORC subunits, these genes are used to express the ORC polypeptides and reconstitute
30 activity. In this way, large amounts of ORC protein from any fungi or mammalian species, including human cells, are obtained.

Inhibitors of ORC protein in fungi provide valuable reagents to selectively inhibit proliferation of fungal cell division by inhibiting the initiation of DNA

replication. This offers a powerful, selective target for antifungal agents valuable in controlling fungal infections in human and other species. For example, as disclosed herein, inhibiting the ORC function by mutation in *S. cerevisiae* can actually cause the death of the mutant cells.

5 In human proliferative disorders such as cancer, cells of the diseased tissue undergo uncontrolled cell proliferation. A key event in this cell proliferation is the initiation of DNA replication. Inhibiting the initiation of DNA replication through inhibition of ORC function provides a valuable target for inhibitors of cell growth. By expressing each of the cDNAs encoding the ORC proteins, either individually
10 or together in an expression system, ORC function is reconstituted *in vitro*. Using this recombinant, expressed protein, inhibitors of ORC function are obtained that block the initiation of DNA replication in cell cycle. As described above, small molecular inhibitors of ORC DNA binding or other activities provide valuable reagents as anti-cancer and anti-proliferation drugs.

15 The following examples are offered by way of illustration and not by way of limitation.

EXAMPLES

Example 1.

20 Transcriptional silencing and ORC.

 The binding of purified ORC to the ARS consensus sequence (ACS) at each of the mating type silencers was tested using a DNase I protection assay (22). ORC protected the match to the ACS at each of the four silencers in an ATP dependent manner. In addition, at each silencer characteristic hypersensitive sites
25 of DNase I cleavage were observed initiating 12-13 bp from the ACS and extending away from the consensus sequence at approximately 10 bp intervals. This pattern of DNase I protection and enhanced cleavage is nearly identical to that observed at non-silencer sequences and indicates that ORC binding to these elements is not fundamentally different from its binding at other ARS elements.

30 At HML-E, HML-I, and HMR-E the only protection observed included the ACS. At HMR-I, however, we observed a second unexpected footprint that did not overlap a strong match to the ACS. Moreover, unlike all previous sites bound by ORC, this protection showed little dependence upon the addition of ATP to the

binding reaction. Although there are two partial matches to the ACS in this region, similar sequences in other ARS elements and silencers were not recognized by ORC, suggesting that these sequences did not direct this unusual ATP-independent binding of ORC to DNA. In combination with the protection observed at the ACS, the boundaries of the ORC footprint at HMR-I were very similar to the boundaries of HMR-I defined by deletion mutagenesis (23). These experiments demonstrate that ORC binds all four of the mating-type silencers, that ORC can bind sequences other than the ACS and that it plays an important role at *HML* and *HMR*.

10 A clear link between ORC function and transcriptional silencing was provided by the finding that a mutation in a gene encoding a subunit of ORC was defective for repression at *HMR* (below). To clone the genes encoding the various ORC subunits, peptides derived from each of the ORC subunits were sequenced (24). A candidate gene, referred to as *ORC2*, was isolated by complementation of 15 a temperature sensitive mutation that showed silencing defects at the permissive temperature. Genetic experiments suggested that *ORC2* mediated the silencing function of the ACS at HMR-E, making it a good candidate to encode a subunit of ORC (below). Comparison of the predicted amino acid sequence of *ORC2* showed that all of the peptides derived from the 72 kd subunit of ORC were within the 20 open reading frame of the *ORC2* gene indicating that it encoded the second largest subunit of ORC.

ORC2 mutations alter ORC function in vitro.

To address the effect of *ORC2* mutations on ORC function in vitro, extracts were prepared from both *orc2-1* and *ORC2* strains (25). Fractions derived from 25 wild-type cells showed strong ORC DNase I protection over the ACS and B1 elements of ARS1 in DNase I footprinting. In contrast, fractions derived from *orc2-1* cells showed a dramatic reduction in ORC DNA binding activity. The ACS and the B1 element were no longer protected from DNase I cleavage. Only the characteristic enhanced DNase I cleavages in the B domain of ARS1 remained. 30 Mutations that disrupt ORC DNA binding at ARS1 prevented the residual DNA binding observed with the mutant fractions, indicating that this binding required the ACS. The DNA binding defects were also not due to a general inhibition of DNA binding as mixing of mutant and wild type fractions did not reduce binding of the

wild type protein. Incubation of the mutant cells at the non-permissive temperature was not necessary to observe defects in ORC DNA binding, which explains the defect observed in mating-type regulation at the permissive temperature (below).

To investigate the polypeptide composition of ORC derived from *orc2-1* and *ORC2* cells, immuno-blots of these fractions were probed with polyclonal antibodies raised against ORC. 30 μ g of partially purified ORC derived from either JRY3688 (*ORC2*) or JRY3687 (*orc2-1*) was separated on a 10% SDS-polyacrylamide gel and transferred to nitrocellulose. The resulting protein blot was incubated with polyclonal mouse sera raised against the entire ORC complex. This sera detects all but the 50 kd subunit of ORC. Antibody-antigen complexes were detected with horseradish peroxidase conjugated secondary antibodies followed by incubation with a chemiluminescent substrate.

Wild type fractions contained the 120, 72, 62, 56, and 53 kd subunits of ORC in roughly equal quantity. The mutant fractions, however, showed a distinctly different subunit composition. While the amount of the 120 and 56 kd subunits was only slightly reduced relative to the wild type fraction, the amount of the 72, 62, and 53 kd subunits was reduced dramatically. In UV cross-linking experiments the same three subunits are specifically cross-linked to DNA in an ACS and ATP dependent manner, suggesting an important role for one or more of these subunits in ORC DNA binding (15). Thus, the absence of these subunits explains the defects in DNA binding observed in vitro and indicates that the *orc2-1* mutation results in a reduction of ORC stability or a defect in Orc2p also results in reduced DNA binding of an intact ORC complex.

orc2-1 cells are defective for entry into S-phase.

The point in the cell cycle the essential function of ORC2 is performed *in vivo* was investigated using alpha factor and hydroxyurea (HU) as cell cycle landmarks (26). Our results were consistent with the execution of the essential function of Orc2p between late G1 and the initiation of DNA synthesis. Arrest with HU followed by release into the non-permissive temperature resulted in 89% of the cells completing an additional cell cycle, indicating that the essential function for Orc2p was executed before the HU arrest point in the cell cycle. In contrast, blocking the cell cycle with alpha-factor followed by release at the non-permissive temperature resulted in the only 41% of the cells completing an additional cell

cycle. This phenotype indicates that the Orc2p function was performed at or near the G1-S phase boundary.

To address the role of ORC in yeast DNA replication more directly, the DNA content of asynchronous cultures of either *orc2-1* or isogenic wild type cells was measured at various times after shifting from the permissive to the non-permissive temperature by fluorescent cytometric analysis (27). JRY3687 (*orc2-1*) or JRY3688 (ORC2) cells grown at 24°C (0 minute time point) or at various times after shifting to the non-permissive temperature (37°C) were fixed, stained with propidium iodide, and analyzed for DNA content using a Coulter Model Epics-C Flow Cytometer. In addition, a small number of cells (approximately 1000) from each time point were returned to the permissive temperature to determine the percentage of cells that remained viable at a given time point. Initially, the DNA content of both wild type and mutant cells was equally divided between 1C and 2C with approximately 10% of the cells in S phase. At early time points after the temperature shift (15-70 minutes) there was a dramatic loss of *orc2-1* cells in S-phase suggesting that entry into S-phase had been halted. Consistent with this hypothesis, as the time course continued the *orc2-1* mutant showed a rapid accumulation of cells with a 1C DNA content and a commensurate decrease in cells with a 2C DNA content (50-100 minutes). Between 100 and 120 minutes, a new population of *orc2-1* cells was observed that appeared to enter into a delayed S phase. By 150 minutes the bulk of the mutant cells were in this population and after 180 minutes only a few cells remained with a 1C DNA content.

Interestingly, we observed a strong correlation between entry into the new round of DNA synthesis and a loss of *orc2-1* cell viability. Similar experiments with isogenic ORC2 cells showed that these effects were specific to the *orc2-1* mutation. These findings indicate that at the non-permissive temperature the *orc2-1* cells were initially unable to enter S phase, but later entered into an abortive round of DNA replication. Entry into this type of replication appears to be a lethal event. Overall, the analysis of the *orc2-1* mutation provides in vivo evidence showing that ORC acts early in S-phase in general, and as the initiator protein at yeast origins of replication in particular.

Identification of the ORC6 gene.

A second gene that represented a strong candidate to encode one of the subunits of ORC was the *AAP1* gene. This gene was cloned using a novel screen for proteins that bound to the ACS in vivo (below). When compared to the predicted amino acid sequence of this gene, we found that all of the peptides
 5 derived from the 50 kd subunit of ORC were encoded by the open reading frame of the *AAP1* gene (28). For this reason we now refer to *AAP1* as *ORC6*, as it encodes the smallest of the six ORC subunits. The identification of this gene as a subunit of ORC provides direct evidence that ORC is bound to the ACS in vivo.

Numbered Citations for Introduction and Example 1

- 10 1. Callan, *Cold Spring Harbor Symp. Quant. Biol.* 38, 195-203 (1973).
2. Fangman and Brewer, *Cell* 71, 363-366 (1992).
3. P. Laurenson and J. Rine, *Micro. Rev.* 56, 543-560 (1992).
4. D. D. Dubey, et al., *Mol. Cell. Biol.* 11, 5346-5355 (1991).
5. D. H. Rivier and J. Rine, *Science* 256, 659-663 (1992).
- 15 6. A. M. Miller and K. A. Nasmyth, *Nature* 312, 247-251 (1984).
7. L. Pillus and J. Rine, *Cell* 59, 637-647 (1989).
8. A. Axlerod and J. Rine, *Mol. Cell. Biol.* 11, 1080-1091 (1991).
9. J. L. Campbell and C. S. Newlon, in *The Molecular and Cellular Biology of the Yeast Saccharomyces* J. R. Broach, J. R. Pringle, E. W. Jones, Eds. (Cold
 20 Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1991) pp. 41-146.
10. J. Broach, et al., *CSH Symp. Quant. Biol.* 47, 1165-1173 (1983).
11. Deshpande and Newlon, *Mol. & Cell. Bio* 12, 4305-4313 (1992).
12. Y. Marahrens and B. Stillman, *Science* 255, 817-823 (1992).
13. S. P. Bell and B. Stillman, *Nature* 357, 128-134 (1992).
- 25 14. Kornberg & Baker, *DNA Replicat'n* (Freeman & Co, NY, 1991) v2.
15. C. S. Newlon, *Microbiol. Rev.* 52, 568-601 (1988).
16. Newlon and Theis, *Current opinion in gen. and dev.* 3, (1993).
17. J. F. X. Diffley and J. H. Cocker, *Nature* 357, 169-172 (1992).
18. Jacob, et al., *CSH Symp. Quant. Biol.*, 28, 329-348 (1963).
- 30 19. DNase I footprinting was performed as previously described (15).
20. J. B. Feldman, et al., *J. Mol. Biol.* 178, 815-834 (1984).
21. To obtain sufficient protein for peptide sequencing, a revised purification procedure for ORC was devised. Whole cell extract was prepared from 400g of

frozen BJ926 cells using a bead beater (Biospec Products) until greater than 90% breakage was achieved. One twelfth volume of a saturated (at 4°C) solution of ammonium sulfate was added to the broken cells and stirred for 30 minutes. This solution was then spun at 13,000 x g for 20 minutes. The resulting supernatant
5 was spun in a 45Ti rotor (Beckman) at 44,000 RPM for 1.5 hrs. 0.27g/ml of ammonium sulfate was added to the resulting supernatant. and the resulting precipitate was collected by spinning in the 45 Ti rotor at 40,000 RPM for 30 minutes. The resulting pellet was resuspended in buffer H/0.0 (15) and dialyzed versus H/0.15M KCl (H with 0.15 M KCl added). Preparation of ORC from this
10 extract was similar to (15) with the following changes. The dsDNA cellulose column was omitted from the preparation and only a single glycerol gradient was performed. Sequencing of peptides derived from ORC subunits was performed using a modification of an "in gel" protocol described previously (40, 41). Purified ORC (~ 10 µg per subunit) was separated by SDS-PAGE and stained with
15 0.1% Coomassie Brilliant Blue G (Aldrich). After destaining the gel was soaked in water for one hour. The protein bands were excised, transferred to a microcentrifuge tube and treated with 200 ng of *Achromobacter* protease I (Lysylendopeptidase: Wako). The resulting peptides were separated by reverse-phase chromatography and sequenced by automated Edman degradation (Applied
20 Biosystems model 470).

22. To isolate and assay ORC from *ORC2* and *orc2-1* cells four liters of JRY3687 (*orc2-1*, *MATa*, *hmrDA::TRP1 ade2 his3 leu2 trp1 ura3*) or the isogenic wild-type strain JRY3688 (*ORC2 MATa*, *hmrDA::TRP1 ade2 his3 leu2 trp1 ura3*) were grown to a density of 2×10^7 cells per ml. Extracts were prepared as
25 described (24) and fractionated over the first two columns in the preparation of ORC. The peak fraction of ORC DNA binding activity eluted from the Q-Sepharose (Pharmacia) column of each preparation was used for subsequent analysis. Antibodies were raised against the entire ORC complex using a single mouse. The resulting sera was able to recognize all but the 50 kd subunit of ORC.
30 Proteins were transferred to nitrocellulose and antigen-antibody complexes were detected with horse radish peroxidase conjugated secondary antibodies and a chemiluminescent substrate.

23. Yeast cells were grown to a density of $1-4 \times 10^7$ cells per ml at 24°C then diluted to a density of $2-4 \times 10^6$ cells per ml into YPD containing $6 \mu\text{M}$ alpha-factor and incubated for 2-2.5 hours at 24°C ($> 90\%$ unbudded cells). For the hydroxyurea arrest experiments alpha factor was washed away and the cells were resuspended in YPD containing 100 mM hydroxyurea and incubated an additional 2.5 hours ($> 90\%$ large budded cells). After incubation with the growth inhibitor, cells were briefly sonicated and plated on YPD plates pre-incubated at either 24°C or 37°C and observed at 0, 3, and 6 hours after plating.
24. Yeast cells were grown to a density of $1-4 \times 10^7$ cells per ml at 24°C and diluted into fresh YPD at either 37°C or 24°C and a density of $2-4 \times 10^6$ cells per ml. At times after dilution, 3×10^6 cells were processed as described (42).
25. The position of the five peptides derived from the 50 kd subunit of ORC in the *ORC6* gene were residues: 51-65; 91-102; 110-105; 207-226; 424-430.
26. K. M. Hennessy, et al, *Genes Dev.* 4, 2252-2263 (1990).
27. H. Renauld, et al., *Genes Dev.* 7, 1133-1145 (1993).
28. A. H. Brand, G. Micklem, and K. Nasmyth, *Cell* 51, 709-719 (1987).
29. McNally and Rine, *Mol. and Cell. Biol.* 11, 5648-5659 (1991).
30. D. D. Brown, *Cell* 37, 359-365 (1984).
31. A. P. Wolffe, *J. Cell Sci.* 99, 201-206 (1991).
32. D. Kitsberg, et al., *Nature* 364, 459-463 (1993).
33. K. S. Hatton, et al., *Mol. Cell. Biol.* 8, 2149-2158 (1988).
34. V. Dhar, et al., *Mol. Cell. Biol.* 9, 3524-3532 (1989).
35. L. G. Edgar and J. D. McGhee, *Cell* 53, 589-599 (1988).
36. L. P. Villarreal, *Micro. Rev.* 55, 512-542 (1991).
37. H. Kawasaki, et al., *Anal. Biochem.* 191, 332-336 (1990).
38. H. Kawasaki and K. Suzuki, *Anal. Biochem.* 186, 264-268 (1990).
39. R. Nash, et al., *EMBL Journal* 7, 4335-4346 (1988).
40. J. Abraham, et al., *J. Mol. Biol.* 176, 307-331 (1984).
41. D. T. Stinchcomb, et al., *Nature* 282, 39-43 (1979).

30

Example 2.

ORC2, a gene required for viability and silencing

In a mutant screen, a temperature-sensitive mutation called *orc2-1* was isolated that, at the permissive temperature, resulted in derepression of *HMR α* flanked by the synthetic silencer and did not cause derepression of *HMR α* flanked by the wild-type silencer (20). Because the *orc2-1* mutant was temperature-sensitive and silencing defective, it merited further analysis. The temperature resistance of a heterozygous *orc2-1/ORC2* diploid (JRY2640) established that the mutation was recessive. The diploid was transformed with a plasmid containing *HMR α* flanked by a mutant silencer (pJR1212), to provide *MAT α 1* function required for sporulation. The temperature-sensitive growth phenotype segregated 2

5 ts : 2 wild type in each of 23 tetrads, indicating that it was caused by a single nuclear mutation. An *HML α mata1 HMR α orc2-1* segregant (JRY3683) was obtained from the diploid following sporulation.

10

Genetic crosses were used to determine which features in the wild-type silencer distinguished it from the synthetic silencer with respect to derepression by *orc2-1*. A *mata1 HMR α* strain (JRY3683) containing the *orc2-1* mutation was mated to a *MAT α* strain containing a mutation in the RAP1 binding site of *HMR-E* flanking *HMR α* (the *HMR α -e-rap1-10* allele; 5401-1a) to determine whether *orc2-1* could derepress *HMR α* in the absence of a functional RAP1 binding site. All 29 of the 96 *MAT α* segregants that had little or no mating ability were temperature-sensitive for growth. Nineteen of the *MAT α* temperature-sensitive segregants were mating competent, indicating that the *orc2-1* mutation *per se* was insufficient to disrupt mating ability, and suggesting that the *HMR α -e-rap1-10* allele was required in combination with *orc2-1* to block mating ability of α strains. A *MAT α* temperature-sensitive segregant from this cross, which mated weakly as an α

20

25 (JRY4133), was confirmed to have the genotype *MAT α HMR α -e-rap1-10 orc2-1*.

As further evidence that *orc2-1* in combination with *HMR α -e-rap1-10* blocked the mating ability of *MAT α* strains, a somewhat unusual cross was used to simplify the previous cross by having *orc2-1* as the only relevant heterozygous marker. Two *MAT α HMR α -e-rap1-10* strains (JRY4133 and JRY4132) had complementary auxotrophic markers, allowing for the selection of the rare *MAT α /MAT α* diploid formed by a mating event between these two strains. This diploid was able to sporulate due to the low level of expression of *HMR α* in the diploid caused by the RAP1-site mutation in the *HMR-E* silencer (21). One of

30

these strains had the *orc2-1* mutation (JRY4133) and the other did not. As expected, the temperature sensitivity segregated 2:2 in each of 34 tetrads. All of the temperature-resistant segregants (two per tetrad) exhibited the α mating phenotype, and all of the temperature-sensitive segregants were either very weak α -matters or were unable to mate at all. The absence of any recombinants between the temperature sensitivity and mating phenotype placed the gene(s) responsible for the temperature sensitivity and the mating defect less than 1.5 centimorgans apart, providing strong evidence that a lesion in a single gene was responsible for both mating phenotypes. This result was in agreement with the co-reversion of the ts and mating phenotypes described herein.

Isolation of multiple alleles of *ORC2*

Using the information from this analysis of *orc2-1*, a second screen was performed to identify additional mutations in essential genes with a role in silencer function. This second screen produced 50 mutants that were temperature sensitive for growth, and in which *HMR α* (flanked by a mutation in the RAP1-binding site) was derepressed at a semi-permissive temperature. Complementation tests for both growth at 37°C and for mating phenotype were performed between *orc2-1* and the collection of temperature-sensitive mutants from the second screen. The collection of temperature sensitive mutants had the *mata1 stel14* genotype, but were able to mate as α 's due to the derepression of *HMR α* . These mutants were mated to a *mata1 orc2-1* strain (JRY3683) and the diploids were tested for growth at 37°C. All but three diploids were able to grow at the restrictive temperature. The three temperature-sensitive diploids were each presumed to be *orc2/orc2* homozygotes due to the inability of the two mutations to complement one another. The mating type of the diploids was checked to determine whether the defect in repression of *HMR* was complemented. All three diploids mated as α 's. Thus, the three mutants were unable to complement either the temperature sensitivity or the mating phenotype of the original *orc2-1* mutation. The new mutations (in strains JRY4136, 4137 and 4138) were designated *orc2-2*, *orc2-3*, and *orc2-4*.

To investigate the possibility that the new mutations were in a gene other than *ORC2* yet still failed to complement *orc2-1*, the allelism between *orc2-1* and *orc2-3* was tested. The original *mata1 orc2-3 stel14* mutant was cured of its *HMR α* plasmid, creating JRY 4137, and mated with a *MAT α HMR α -e-rap1-10*

orc2-1 strain (JRY3685). In 24 tetrads from this diploid, all segregants were temperature sensitive for growth, indicating strong linkage between *orc2-1* and *orc2-3* (<2 centimorgans). All further studies were performed using the *orc2-1* allele, which provided the stronger mutant phenotypes.

5 Map position of *ORC2*

Linkage between *ORC2* and *LYS2*, on chromosome II, was evident in crosses between two *lys2* strains (JRY2640 and PSY152) and the original *orc2-1* isolate (JRY2903) that placed *ORC2* approximately 24 centimorgans from *LYS2*. A third cross (JRY4130 x JRY4134) tested the linkage between *sec18*, which is
 10 centromere proximal to *LYS2*, and *ORC2*. Because both *orc2-1* and *sec18* are temperature sensitive, an *ORC2* allele marked by *URA3* (from pJR1423) was used to determine that *SEC18* and *ORC2* were separated by 6.6 centimorgans (Table 1). No previously-mapped genes involved in silencing map near *SEC18*.

Table 1. Linkage of *ORC2* to *LYS2* and *ORC2* to *SEC18*

15	distance	Tetrad types			Map
		<u>PD</u>	<u>T</u>	<u>NPD</u>	
	<u>Cross</u>				<u>(cM)</u>
	<i>ORC2</i> vs <i>LYS2</i>	10	14	0	29
	<i>ORC2</i> vs <i>LYS2</i>	20	14	0	21
20	<i>ORC2</i> vs <i>LYS2</i> TOTAL	30	28	0	24
	<i>ORC2</i> vs <i>SEC18</i>	46	7	0	6.6

The *ORC2* mutants arrested with a cell cycle terminal phenotype.

The effect of the *orc2-1* mutation on the cell division cycle was explored:
 25 mutant *orc2-1* strains were grown in liquid medium at 23°C, the permissive temperature, and then shifted to 37°C to test whether the cells arrested with a single terminal morphology. Specifically, *orc2-1* cells (JRY3683) were grown to log phase at the permissive temperature (23°C) and the culture was split. Half of the culture was grown an additional five hours at the permissive temperature and
 30 the other half was shifted to the nonpermissive temperature (37°C) and grown for an additional five hours. At that time, both cultures were fixed and stained with DAPI to allow visualization of the nucleus. In the culture maintained at the permissive temperature, cells at all phases of the cell cycle were observed. Cells

later in the cell cycle, as evidenced by the presence of large buds, frequently exhibited nuclei in both the mother and the daughter cell. In contrast, in the culture shifted to the restrictive temperature, approximately 90% of the cells arrested as large budded cells. Nuclei were only present in the mother cell and not
5 in the daughter cells. In addition, the cells were larger than those grown at the permissive temperature, indicating that protein synthesis and cell wall synthesis continued in the absence of *ORC2* function. Similar results were obtained with two additional *orc2-1* strains (JRY3685 and JRY3687).

ORC2 cells harvested either after continuous growth at the permissive
10 temperature or after a shift to the nonpermissive temperature were fixed and stained with DAPI allowing visualization of DNA with fluorescence microscopy. The cells grown permissively displayed a range of morphologies from small unbudded cells to cells with single buds of various sizes. The cells shifted to the nonpermissive temperature looked very different: the majority arrested as large
15 budded cells, and for the most part, each mother-daughter pair contained only a single brightly-staining region, often at or near the neck. These data indicated that *orc2-1* mutants displayed cell cycle defects characteristic of mutants defective in DNA replication.

20 Cloning of the *ORC2* gene:

The *ORC2* gene was cloned by complementation of the *orc2-1* temperature sensitivity (22). One complementing clone (pJR1416) was chosen for further analysis. Subclones missing various fragments from the insert were retransformed into an *orc2* strain to assay whether the deletion affected the clone's ability to
25 complement *orc2-1*'s temperature sensitivity. The key observations were that the deletion of a 2.8-kb *Sst*I-*Sst*I fragment destroyed complementation activity, whereas the deletions of flanking sequences (*Xba*I, and the larger *Sst*I fragment) had no effect. The 2.8-kb fragment was subcloned (pJR1263), and shown to possess complementing activity.

30 To determine whether the gene on the clone was indeed allelic to the *ORC2* mutation, a fragment of the original clone was subcloned into a yeast integrating vector. This plasmid (pJR1423) was cleaved within the insert to direct homologous integration and transformed into a wild-type strain (W303-1A). As a result, the

site of integration was marked by the plasmid's *URA3* gene. The resulting strain (JRY4134) was crossed to an *orc2-1* strain (JRY3685). In each of 59 tetrads, *URA3* segregated opposite to the temperature sensitivity caused by *orc2-1*, indicating that *ORC2* had indeed been cloned.

5 *ORC2* was essential for cell viability.

ORC2 was disrupted by *URA3*, (23), and integrated into a diploid homozygous for *ura3* and *ORC2*, (JRY3444). Of the 41 tetrads dissected, 40 tetrads had two live and two dead segregants, and one tetrad had only one live segregant. The colonies that grew were, without exception, Ura-. By inference, 10 the dead segregants contained the *URA3* gene, and thus the *ORC2* disruption, indicating that *ORC2* function was essential for cell viability at all temperatures. The dead segregants were examined under a microscope to gain some insight into the true null phenotype. Most of the spores germinated into cells that were elongated or otherwise deformed and had not divided. In no case did the cell 15 divide more than two times. Thus in many spores, the absence of *ORC2* blocked cell division but not growth.

Role of *ORC2* in Plasmid Replication

 To test the role of *ORC2* in plasmid stability, an isogenic pair of strains, one wild type (W303-1B) and one *orc2-1* (JRY4125), were transformed with a 20 plasmid containing a centromere, a suppressor tRNA (*SUP11-1*), *URA3*, and *ARS1*, a chromosomal origin of replication (*YRP14/CEN4/ARS1/ARS1*; (24, 25), selecting for uracil prototrophy. Transformants were grown on selective medium at 23°C, the permissive temperature for *orc2-1*. The colonies were picked from the selective plate, serially diluted, plated onto solid rich medium and grown to 25 colonies at 23°C. The wild-type transformants grew into colonies most of which were white with a few exhibiting red sectors. The small fraction of red colonies were from cells in the selectively grown colony that had lost the plasmid. In contrast, the majority of colonies from the *orc2-1* mutant were red, reflecting a high degree of plasmid loss among the cells in the selectively grown colony. 30 Moreover, in the *orc2-1* strain, red sectors were present in the majority of white colonies with some white colonies displaying multiple red sectors.

 It is possible to quantitate the number of cell cycles in which a plasmid is lost from the number of colonies that are half red and half white. Only those

colonies that lose the plasmid in the first cell division form half red, half white colonies. In the case of the wild-type strain, 0.9 % (10 / 1168) of the colonies were half red and half white, indicating that the plasmid was lost in 0.9 % of cell cycles. In contrast, the frequency of half red and half white colonies in the *orc2-1* strain grown at the permissive temperature was 11 % (58 / 512), indicating that the same plasmid was lost approximately 12 times as often in the strain with partially defective Orc2p. These data indicated a profound defect in plasmid stability specific to the *orc2-1* strain, and in combination with the cell-cycle phenotype of *orc2-1*, suggested that *orc2-1* strains were defective in DNA replication. These results were consistent with the flow cytometry studies of *orc2-1* strains herein.

Sequence of *ORC2*

The sequence of the 2.8-kb *SsrI-SsrI* *orc2*-complementing fragment was determined and deposited in Genbank (Accession #L23924). The only open reading frame of significant length was deduced to be *ORC2*, and predicted a 620 residue protein of approximately 68 kD. The *SsrI* fragment included 806-bp of upstream sequence and 140-bp of downstream sequence.

The deduced Orc2p protein was 15% basic residues and 16% serine/threonines. Fully 50% of the N-terminal residues (residues 15-280) were lysine, arginine, proline, serine, or threonine. The KeyBank motif program revealed several matches to peptide motifs within Orc2p. Orc2p contained many potential phosphorylation sites: 3 for cAMP- and cGMP-dependent protein kinase (starting at residues 57, 433 and 546), 12 for protein kinase C (24, 41, 42, 89, 101, 102, 176, 321, 335, 431, 521, and 549) and 14 for casein kinase II (60, 148, 149, 182, 238, 270, 389, 481, 486, 491, 505, 552, 595, and 605), and match to the nuclear targeting sequence (residues 103-107). A perfect match to the RAP1 binding site consensus (starting at nucleotide 595), and two near matches (12/15) to the ABF1-binding consensus sequence (starting at 12 and 609). It was determined by sequence homology that a lysyl tRNA synthetase gene is located to the left of the *SsrI* fragment shown here (Mirande and Waller, 1988), and a kinase homolog to the right.

Another homology is with the region near the catalytic domain of human topoisomerase I proteins which has diverged among topoisomerase I proteins from other species except for the region surrounding the invariant active-site tyrosine.

This region includes a consensus sequence consisting of a serine and lysine residue near the tyrosine (25). The Orc2p protein also contained such a consensus sequence near its C-terminus. However, mutation of this putative active-site tyrosine to phenylalanine had no detectable effect on the ability of *ORC2* to complement the temperature-sensitivity or mating defect of an *orc2-1* strain.

Table 2. Strain list.

Strain	Genotype ^(a)
DBY1034	<i>MATa his4-539 lys2-801 ura3-52 SUC</i>
W303-1A	<i>MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1</i>
10	<i>ura3-1</i>
W303-1B	<i>MATα ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1</i>
	<i>ura3-1</i>
PSY152	<i>MATa his3D200 leu2-3,112 lys2-801 ura3-52</i>
JRY4130	<i>MATα his4 ura3 sec18</i>
15	JRY438 <i>MATα Gal⁺ his4-519 leu2-3,112 SUC2 ura3-52</i>
	JRY543 <i>MATa/MATα ade2-101/ade2-101 his3Δ200/his3Δ200</i>
	<i>lys2-801/lys2-801 met2/MET2 TYR1/tyr1</i>
	<i>ura3-52/ura3-52</i>
	JRY2640 <i>mata1 ade2 leu2-3,112 lys2-801 ura3</i>
20	JRY2698 <i>MATα HMRα ade2-101 his3 leu2 trp1 ura3-52</i>
	JRY2699 <i>MATα HMRα ade2-101 his3 leu2 trp1 ura3-52</i>
	<i>sir4DN::HIS3</i>
	JRY2700 <i>MATα HMRα ade2-101 his3 leu2 trp1 ura3-52</i>
	+ pJR924
25	JRY2903 <i>MATα HMRα ade2-101 his3 leu2 orc2-1 trp1 ura3-52</i>
	JRY2904 <i>MATα HMRα ade2-101 his3 leu2 orc2-1 trp1 ura3-52</i>
	+ pJR924
	JRY3444 <i>MATa/MATα ade2-101/ade2-101 his3D200/his3D200</i>
	<i>lys2-801/lys2-801 met2/MET2 TYR1/tyr1</i>
30	<i>ura3-52/ura3-52 orc2::Tn10LUK/ORC2</i>
	JRY3683 <i>mata1 {HMRα} ade2 his3 leu2 orc2-1 ura3</i>
	JRY3685 <i>MATα HMRA-e-rap1-10 ade2 leu2 trp1 orc2-1 ura3</i>
	JRY3687 <i>MATα hmrDA::TRP1 ade2 his3 leu2 trp1 ura3 orc2-1</i>

- JRY3690 *MATa HMRe-e-rap1-10 ade2 his3-11,15 leu2 orc2-1
trp1 ura3*
- JRY4125 *MATa ade2-1 can1-100 his3-11,15 leu2-3,112 orc2-1
trp1-1 ura3-1*
- 5 JRY4132 *MAT α HMRe-e-rap1-10 ade2 his3 ura3*
- JRY4133 *MAT α HMRe-e-rap1-10 ade2 leu2 orc2-1trp1 ura3*
- JRY4134 *MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1
ura3-1 ORC2::pJR1423*
- JRY4135 *matal ade2 leu2-3,112 lys2-801 ura3 stel4*
- 10 JRY4136 *matal ade2 leu2-3,112 lys2-801 orc2-2 ura3 stel4*
- JRY4137 *matal ade2 leu2-3,112 lys2-801 orc2-3 ura3 stel4*
- JRY4138 *matal ade2 leu2-3,112 lys2-801 orc2-4 ura3 stel4*

(a) Unless otherwise noted, all strains were *HML α* and *HMRe*. *HMRe-e-rap1-10* refers to the allele of *HMR-E*, originally described as *PAS1-1*, that contains a mutation in the RAP1 binding site (21).

Numbered Citations for Example 2.

1. I. Herskowitz, et al *Cold Spring Harbor Laboratory Press* 583 (1992).
- 20 2. J. Abraham, J. Feldman, K.A. Nasmyth, J.N. Strathern, J.R. Broach, and J. Hicks, *C.S.H. Symp. Quant. Biol.* 47, 989 (1982). J.B. Feldman, J.B. Hicks, and J.R. Broach, *J. Mol. Biol.* 178, 815 (1984).
3. J. Rine, and I. Herskowitz, *Genetics* 116, 9 (1987).
4. Kurtz et al, *Genes Dev.* 5, 616 (1991); Sussel et al, *PNAS* 88, 7749 (1991).
- 25 5. J.R. Mullen, et al, *PMBO J.* 8, 2067 (1989).
6. P.S. Kayne, et al, *Cell* 55, 27 (1988). L.M. Johnson, et al, *Proc. Natl. Acad. Sci. USA* 87, 6286 (1990). P.D. Megee, et al, *Science* 247, 841 (1990). E. Park, and J. Szostak, *Mol. Cell. Biol.* 10, 4932 (1990).
7. P. Laurenson, and J. Rine, *Microbiol. Rev.* 56, 543 (1992).
- 30 8. Brand, et al., *Cell* 41, 41 (1985); Kimmerly, et al., *EMBO J.* 7, 2241 (1988).
9. D. Shore, and K. Nasmyth, *Cell* 51, 721 (1987).

10. M.S. Longtine, et al., *Curr. Genet.* 16, 225 (1989).
11. A.R. Buchman, et al, *Mol. Cell. Biol.* 8, 5086 (1988).
12. J.F.X. Diffley, and J.H. Cocker, *Science* 357, 169 (1992).
13. A.S. Buchman, and R.D. Kornberg, *Mol. Cell. Biol.* 10, 887 (1990).
- 5 14. J.A. Huberman, et al, *Nucleic Acids Res.* 16, 6373 (1988). B.J. Brewer, and W.L. Fangman, *Cell* 51, 463 (1987).
15. S.P. Bell and B. Stillman, *Nature* 357, 128 (1992).
16. F.J. McNally, and J. Rine, *Mol. Cell. Biol.* 11, 5648 (1991).
17. A.M. Miller, and K.A. Nasmyth, *Nature* 312, 247 (1984).
- 10 18. D.H. Rivier, and J. Rine, *J. Science* 256, 659 (1992).
19. Two genetic screens were devised to identify temperature sensitive mutations in essential genes involved in silencing. The screen that led to isolation of *orc2-1* started with JRY2698 (*HML α* , *MAT α* , *HMR α* , *ade2*, *his3*, *leu2*, *trp1*, *ura3-52*), which had a mating-type cassettes at all three chromosomal mating-type
- 15 loci and was transformed with a plasmid (pJR924) containing the a mating-type cassette at *HMR* (JRY2700). The plasmid-borne *HMR α* locus had two synthetic silencers substituted for the *E* silencer, and also had a deletion of the *I* element. The use of two silencers rather than one minimized the risk of being distracted by site mutations in the silencer. One hundred and sixty two thousand colonies of
- 20 EMS-mutagenized colonies were grown on supplemented minimal media (without uracil) at 25°C and screened for derepression of the plasmid-borne a cassette at *HMR*. Mutagenized colonies were replica-plated onto lawns of the mating tester strain DBY1034 (*MAT α* , *his4-539*, *lys2-801*, *ura3-52*) on minimal media either with or without uracil supplementation. Replicas were incubated at 25°C for one
- 25 hour, then overnight at 30°C. Only plasmid-containing JRY2700 cells were able to mate with the tester strain to yield diploids capable of growing on the unsupplemented plates because the only functional *URA3* gene was on the plasmid.

Cells bearing mutations causing derepression of the plasmid-borne a cassette could be distinguished from the other classes of mutations by exploiting a

30 feature of yeast plasmids. Approximately 10% of the cells in these colonies lacked the plasmid and thus could, in principle, mate with the tester strain and form Ura diploids capable of growth on the plates supplemented with uracil. If a colony had a mutation in the mating response pathway, the cells would be unable to mate even

- in the absence of the plasmid, and thus would be unable to form diploids capable of growth on medium supplemented with uracil. Twenty eight strains were identified that were temperature-sensitive for growth and that mated with the tester strain only on plates supplemented with uracil. Plasmid-free isolates of each strain
- 5 were then retransformed with the plasmid bearing the synthetic silencer at the *HMRa* locus (pJR924) and with the plasmid bearing the wild-type *HMRa* locus (pJR919; McNally and Rine, 1991). Three strains were able to mate when carrying the wild-type *HMR* locus (pJR919) but not when carrying the synthetic silencer-containing *HMR* locus (pJR924). In order to determine if the ts growth
- 10 phenotype and the mating phenotype were due to the same mutation, spontaneous revertants of the ts phenotype were selected. A spontaneous revertant of the ts growth of one strain, JRY2904, mated as well as the wild-type JRY2700, suggesting that the mating phenotype and temperature-sensitive growth were due to the same mutation which was named *orc2-1*.
- 15 20. Y. Kassir, et al, *Genet.* 109, 481 (1985). Foss and Rine, *Genetics.* (1993)
21. The *ORC2* gene was cloned by complementation of the temperature sensitivity of *orc2-1*. An *orc2-1* strain (JRY3683) was transformed with a CEN *LEU2*-based *Saccharomyces cerevisiae* genomic library (32) Approximately 1000 to 1500 transformants formed colonies at 23°C. Replica prints of these colonies
- 20 were incubated at 37°C to screen for the ability to grow at elevated temperatures. Plasmids were isolated from temperature-resistant strains and retested. Those plasmids that complemented the defect a second time were analyzed by restriction digestion. One plasmid from the CEN-*LEU2* library (pJR1416) was chosen for further analysis.
- 25 22. *ORC2* was disrupted with the *Tn10* LUK transposon (33), which inserted within the *ORC2* coding sequence on the plasmid (pJR1146) carrying the *SstI* *orc2-1* complementing fragment. Plasmid pJR1147 had the *Tn10*LUK insertion within the *ORC2* coding region. The *ORC2*-containing *SstI* fragment, disrupted by the transposon, was removed from pJR1147 by partial digestion with *SstI*. The
- 30 fragment was transformed into the wild-type diploid JRY543. The integration of this disruption allele at the *ORC2* locus was confirmed by DNA blot hybridization analysis (Southern, 1975), and the diploid was named JRY3444.
23. P. Hieter, C. Mann, M. Snyder, and R.W. Davis, *Cell* 40, 381 (1985).

24. D. Koshland, J.C. Kent, and L.H. Hartwell, *Cell* 40, 393 (1985). R.M. Lynn, et al, *Proc. Natl. Acad. Sci. USA* 86, 3559 (1989). W.-K. Eng, S.D. Pandit, and R. Sternglanz, *J. Biol. Chem.* 264, 13373 (1989).
- 25, 26. A.H. Brand, G. Micklem, and K. Nasmyth, *Cell* 51, 709 (1987).
- 5 27. S. Shuman, et al, *Proc. Natl. Acad. Sci. USA* 86, 9793 (1989).
- 28, 29. J. Singh, and A.J.S. Klar, A. J. S. *Genes and Dev.* 6, 186 (1992).
30. D.D. Dubey, et al, *Mol. Cell. Biol.* 11, 5346 (1991).
31. C.A. Hrycyna, et al, *EMBO J.* 10, 1699 (1991).
32. A mutation was introduced into the RAP1 binding site at *HMR-E* adjacent
10 to the *HMR α* locus by oligonucleotide-directed mutagenesis (35), and the change confirmed by sequencing. The RAP1 site mutation was identical to the *PAS1-1* mutation of *HMR-E* characterized previously that blocks RAP1 protein binding in vitro (21), and is described here as *HMR α -e-rap1-10*. The plasmid consisting of the *HMR α -e-rap1-10* *HindIII* fragment in pRS316 was named pJR1425. The wild-
15 type *HMR α* version of the same plasmid was named pJR1426. Approximately 100,000 mutagenized cells from 12 independent cultures of the *HML α mata1 HMR α ste14* strain with the *HMR α* plasmid (pJR1425) were grown into colonies at 23°C and replica-plated to a *MAT α ura3* mating-type tester lawn (PSY152) to identify mutants exhibiting the a mating phenotype. The mating plates were
20 incubated at 30°C in order to identify mutants defective enough to be derepressed at *HMR* yet not so defective as to be inviable. Of nine hundred haploid mating proficient colonies that were picked, fifty mutants were temperature sensitive for growth at 37°F to some degree. These mutants were subjected to further study and the remainder were discarded. All 50 mutants were recessive to wild-type. Only
25 the subset of mutants relevant to *ORC2* are presented here; the remainder will be discussed elsewhere.
33. The *ORC2* gene was defined by the *orc2-1* mutation. An *orc2*-complementing plasmid (pJR1416) was obtained by complementation of the temperature sensitivity of *orc2-1*. In order to map the approximate position of the
30 *orc2* -complementing gene in the plasmid, six derivatives of pJR1416 were made and tested for complementation. The *SalI-SalI* fragment was removed from the insert to yield pJR1418. Three adjacent *XbaI-XbaI* fragments were removed to

yield pJR1422. *SphI* cleaved once in the insert and once just inside the vector. Deleting this *SphI-SphI* fragment produced pJR1417. Cleavage by *SstI* released two fragments from the insert. Deletion of both fragments created pJR1419. Isolates in which only the larger *SstI* fragment (pJR1421) or only the smaller *SstI* fragment (pJR1420) was deleted were also recovered. The 2.8-kb *SstI-SstI* *orc2*-complementing fragment was cloned into the *SstI* site of the CEN *URA3* vector pRS316 (36), to yield pJR1263. Two plasmids were made which allowed the chromosomal integration of part or all of *ORC2*. The first, pJR1423, contained an *XhoI/KpnI* insert (from pJR1416) which extended from a few kb upstream of the *ORC2* start codon to about 60-bp upstream of the stop codon inserted into *XhoI-KpnI*-cut pRS306 (36), a yeast integrating vector marked by *URA3*. The second plasmid, pJR1424, contained the *SstI* *orc2*-complementing fragment inserted into the *SstI* site of pRS306.

34. F. Spencer, et al *Genetics* 124, 237 (1990).
35. O. Huisman, et al, *Genetics* 116, 191 (1987).
36. E.M. Southern, *J. Mol. Biol.* 98, 503 (1975).
37. T.A. Kunkel, et al, *Methods Enzymol.* 154, 367 (1987).
38. R.S. Sikorski, and P. Hieter, *Genetics* 122, 19 (1989).

Example 3.

In order to identify potential yeast initiators, we developed a genetic strategy, the one-hybrid system, to find proteins that recognize a target sequence of interest. The one-hybrid system has two basic components: (i) a hybrid expression library, constructed by fusing a transcriptional activation domain to random protein segments, and (ii) a reporter gene containing a binding site of interest in its promoter region. Hybrid proteins that recognize this site are expected to induce expression of the reporter gene, because of their dual ability to bind the promoter region and activate transcription (8). This association may be indirect, since hybrids that interact with endogenous proteins already occupying the binding site will also activate transcription (7). Nevertheless, as long as the association is sequence-specific the protein incorporated in the hybrid should be functionally relevant.

We have used this method to look for proteins from the yeast *Saccharomyces cerevisiae* that recognize the *ARS* consensus sequence (ACS) of yeast origins of DNA replication. The protein component of this screen was provided by a set of three complementary yeast hybrid expression libraries, YL1-3, 5 containing random yeast protein segments fused to the GAL4 transcriptional activation domain (GAL4^{AD}) (9). The reporter gene for our screen contained four direct repeats of the ACS in its promoter region and was integrated into the yeast strain GGY1 to form JLY363(ACS^{WT}) (10). To determine the dependence of *lacZ* induction on the ACS, we constructed in parallel JLY365(ACS^{MUTANT}), which 10 harbors a reporter gene carrying four copies of a nonfunctional multiply-mutated ACS (Fig. 4) (10).

We isolated nine plasmids that induced greater *lacZ* activity in JLY363(ACS^{WT}) than JLY365(ACS^{MUTANT}) from a screen of 1.2 million YL1-3 transformants (11). Many of the plasmids that induced *lacZ* activity on initial 15 screening of the library in JLY363(ACS^{WT}) failed to exhibit a dependence on the ACS when introduced into JLY365(ACS^{MUTANT}). Restriction analysis of these plasmids showed that the nine isolates represented five genomic clones, which we initially labeled *AAP1-5* for *ACS* associated protein. *AAP1* was isolated four times, *AAP5* twice, and the others only once.

20 To examine the sequence specificity of *lacZ* induction with finer resolution, reporter constructs containing direct repeats of four ACS point mutants were each integrated into GGY1 to generate the set of reporter strains(10). The five *AAP* clones were individually examined in these strains for the ability to induce *lacZ* expression. *AAP1* displayed a correspondence between the induction of this set of 25 reporter genes and the *ARS* function (12) of their ACS. The *AAP5* hybrid exhibited a slightly weaker correlation, and the remaining clones showed poor correlation. These findings suggest that *AAP1*, and possibly *AAP5*, encodes a protein that recognizes the ACS in a sequence-specific manner. Constructs with deletions in the *AAP1* coding sequence (14) were unable to induce *lacZ* expression, 30 indicating that recognition of the ACS resided in the protein segment fused to GAL4.

The genomic segments fused to the GAL4^{AD} in *AAP1-5* were sequenced (15) to determine the extent of the hybrid proteins that were made. *AAP1* and *AAP5*

had sizable protein coding sequences of 301 and 123 amino acids, respectively, fused in frame with the GAL4^{AD}. In principle, these segments are large enough to direct the hybrid protein to the promoter of the reporter gene. *AAP2-4* encoded hybrid proteins with only short peptide extensions (10, 22, and 38 amino acids
5 respectively) fused to the GAL4^{AD}, suggesting that these hybrids were not responsible for the transcriptional induction attributed to these clones. Because of this finding and the lack of proper sequence specificity for the ACS element, *AAP2-4* were not studied further.

The full-length gene for *AAP1* was cloned from a yeast genomic library
10 and sequenced (15) (Genbank accession no. L23323). *AAP1* contains an open reading frame for a protein 435 amino acids long with a predicted molecular weight of 50,302 daltons. The hybrid GAL4^{AD}-*AAP1* protein obtained from the screen was a fusion of the GAL4^{AD} to the C-terminal two-thirds of the predicted full-length protein (residues 135-435), indicating that this portion of the molecule
15 is sufficient for association with the ACS. Comparison of peptide sequences from the 50kd subunit of ORC with the predicted protein sequence from *AAP1* demonstrated that our gene encodes this subunit and confirmed the association between the *AAP1* protein and the ACS. Because of this identity, we have renamed our gene *ORC6*.

20 An overlapping ORF capable of encoding a protein 250 amino acids long exists on the complementary strand. The positions of the predicted start and stop codons for this ORF are at nt 1615-7 and nt 865-7, respectively. In pJL766 the C residue at 1471 was mutated to a T, preserving the amino acid sequence of *ORC6* but introducing a stop codon in this overlapping ORF. The sequence of *ORC6*
25 indicates a connection with the regulatory machinery governing cell cycle progression. *Orc6p* contains four phosphorylation sites, (S/T)PXX, for cyclin-dependent protein kinases (20) clustered in the first half of the molecule. Using the more relaxed consensus site (S/T)P adds two more sites to this cluster. We have observed *Orc6p* phosphorylated *in vivo* on serine and threonine residues.
30 However, since the initiation of yeast DNA replication commences promptly in response to the activation of this protein kinase in G1, we believe that *Orc6p* and possibly other ORC subunits are regulated substrates of this kinase. Finally, as expected for a protein participating in nuclear events, *Orc6p* contains a potential

nuclear localization signal (NLS) within the (S/T)PXK cluster and one in the C-terminal domain (amino acid residues 117-122 and 263-279). Orc6p can be seen in the nucleus by immunofluorescence.

A marked deletion of the *ORC6* gene (pJL731) (21), removing all but 13
5 codons from its open reading frame, was introduced into diploids from three different strain backgrounds. The resulting heterozygous *ORC6* deletion strains, JLY481, JLY475, and JLY469 were induced to undergo meiosis, and 20 tetrads of each strain were dissected (21). In all backgrounds the *ORC6* disruption cosegregated with inviability, demonstrating that *ORC6* is essential for cell growth.
10 Microscopic examination revealed that mutant spores from JLY481 and JLY475 germinated, completed 1-2 rounds of cell division, and then arrested with a uniform large bud morphology reminiscent of cell division cycle mutants defective in DNA replication or nuclear division (22). The position of cell cycle arrest could not be pinpointed, however, since the DNA content of these cells could not be
15 readily measured. Mutant spores derived from JLY469 germinated poorly.

The interpretation of these *ORC6* deletion experiments was complicated by the presence of a second open reading frame (ORF2) of 250 amino acids on the antisense strand of the *ORC6* gene. ORF2 spans nucleotides 1617 to 868 of the Genbank sequence and overlaps the C-terminal two-thirds of the *ORC6* coding
20 sequence. A marked deletion that removed the N-terminal third of the *ORC6* coding sequence without affecting ORF2 (pJL733) was introduced into diploids (21). Tetrad analysis again showed the *ORC6* deletion cosegregating with cell death. Finally, an *ORC6* gene was constructed that contains a silent codon change for the *ORC6* ORF but introduces a UGA stop codon in ORF2 (22). This gene
25 was able to rescue a haploid strain containing a full deletion of the *ORC6* ORF. We conclude that *ORC6* is essential for cell viability.

Our results validate the one-hybrid system screen as a method to identify and clone genes for proteins that recognize a DNA sequence of interest. This screen has also been successful in identifying DNA-binding proteins (23), and a
30 variation of this screen has been used to identify a binding site for a suspected DNA-binding protein (24). The one-hybrid approach is particularly useful for proteins that are difficult to detect biochemically or for which starting material in a purification is difficult to obtain.

We identified genes that interact genetically with *ORC6* using established *cdc* mutants because germinating spores bearing an *ORC6* deletion appeared to exhibit a cell division cycle phenotype. pJL749 (28), a plasmid that overexpresses Orc6p several hundred-fold, was introduced into a virtually isogenic set of temperature-sensitive *cdc* mutants arresting at various points in the cell cycle (29). Overexpression of *ORC6* selectively affected *cdc6* and *cdc46* mutants, lowering their restrictive temperature by 5-7° C; there was no significant effect on the other mutants examined or on the wild-type strain (Table 1).

10	Strain	cdc mutant	viability with overexpression of ORC6
	RDY488	wild-type	+
	RDY501	cdc28-1	+
	RDY510	cdc4-1	+
	RDY664	cdc34-2	+
15	RDY543	cdc7-4	+
	JLY310	cdc6-1	-
	JLY179	cdc46-1	-
	JLY338	cdc2-1	+
	JLY353	cdc17-1	+
20	RDY619	cdc15-2	+

Table 1. Viability of *cdc* Mutants in the Presence of High Levels of *ORC6* Expression. JL749 (GALp-HA-ORC6), JL772 (GALp-HA), and RS425 were introduced into each *cdc* mutant, and examined for growth at various temperatures under conditions that induce expression of ORC6 (28, 29). + indicates mutants whose restrictive temperature remains unchanged in the presence of JL749 relative to JL772 and RS425. - indicates mutants whose restrictive temperature is lowered 5-7° C when JL749 is present.

Numbered Citations for Example 3

1. Kelly, J. *Biol. Chem.* 263, 17889 (1988); Mariani, *Annu. Rev. Biochem.* 61, 673 (1992); Kornberg, Baker, *DNA Replication*. (Freeman and Company, New York, 1992); B. Stillman, *Annu. Rev. Cell Biol.* 5, 197 (1989).
- 5 2. M. L. DePamphilis, *Annu. Rev. Biochem.* 62, 29 (1993).
3. Campbell and Newlon, in *The Molecular and Cellular Biology of the Yeast Saccharomyces* Broach, et al, Eds. (CSHL Press, 1991), vol. 1, pp. 41-146.
4. Fangman and Brewer, *Annu. Rev. Cell Biol.* 7, 375 (1991).
5. J.R. Broach et al., *Cold Spring Harbor Symp. Quant. Biol.* 47, 1165
- 10 (1983); Van Houton and C. S. Newlon, *Mol. Cell. Biol.* 10, 3917 (1990).
6. Y. Marahrens and B. Stillman, *Science* 255, 817 (1992).
7. S. Fields and O.-K. Song, *Nature* 340, 245 (1989); C.-T. Chien, P.T. Bartel, R. Sternglanz, S. Fields, *Proc. Natl. Acad. Sci. USA* 88, 9578 (1991).
8. R. Brent and M. Ptashne, *Cell* 43, 729 (1985).
- 15 9. The N-terminal portions of the hybrids from three related hybrid expression libraries, YL1-3 (7), consist of the SV40 nuclear localization signal and amino acids 768-881 of the GAL4 activation domain (GAL4^{AD}). The C-terminal portions were derived from random yeast protein segments which have been fused to the end of the GAL4^{AD}. These segments are encoded by short (1-3kb) fragments from
- 20 a Sau3a partial digest of yeast genomic DNA. Together, YL1-3 ensure that all three reading frames of these fragments can be expressed.
10. pLR1D1 is described in R.W. West Jr., R.R. Rogers, M. Ptashne, *Mol. Cell. Biol.* 4, 2467 (1984). We generated pBgl-lacZ from pLR1D1 by (i) substituting an XhoI-BglII-XhoI polylinker for the XhoI linker and (ii) precisely
- 25 excising a Hind III fragment containing 2m sequences. The resulting vector has a unique Bgl II site approximately 100 bp upstream of the TATA box for insertion of DNA sequences in the promoter region and a unique StuI site for targeted integration of the plasmid at the URA3 locus. Multiple direct repeats of ARS1 domain A and several of its mutant derivatives were inserted into the Bgl II site of
- 30 pBgl-lacZ to generate all the reporter genes used in this work. The inserted repeat elements, derived from complementary oligonucleotides, were oriented with the TATA box to their right. Each reporter gene construct was integrated into the

URA3 locus of GGY1 (*MATa Dgal4 Dgal80 ura3 leu2 his3 ade2 tyr*) [G. Gill and M. Ptashne, *Cell* 51, 121 (1987)] to create a reporter strain. Integration of pBgl-lacZ into GGY1 generated JLY387.

11. YEPD (rich complete) and SD (synthetic dropout) media are as described
5 [J.B. Hicks and I. Herskowitz, *Genetics* 83, 245 (1976)]. Standard methods were used for manipulation of yeast cells [C. Guthrie and G.R. Fink, Ed., *Guide to Yeast Genetics and Molecular Biology* (Academic Press, San Diego 1991)] and DNA [F.M. Ausubel et al., Ed., *Current Protocols in Molecular Biology* (Wiley, New York 1989)]. Libraries YL1-3 were transformed [R.H. Schiestl and R.D.
10 Geitz, *Current Genetics* 16, 339 (1989)] into JLY363 (10) and plated on SD-Leu at a density of 2-5000 colonies/10cm plate. 500,000 transformants were obtained for YL1 and YL2, and 200,000 for YL3. Transformants were assayed on filters for production of b-galactosidase [L. Breeden and K. Nasmyth, *Cold Spring Harbor Symp. Quant. Biol.* 47, 643 (1985)]. 49 isolates remained positive after colony
15 purification (15 from YL-1; 22 from YL-2, 12 from YL-3), and library plasmids were extracted from them. These plasmids were each transformed into both JLY363 and its mutant counterpart JLY365 (10). Nine plasmids induced greater b-galactosidase activity in the wild type reporter strain than the control. These plasmids were classified into five clones, *AAP1-5*, based on their Hind III
20 restriction pattern. Each clone was then retested in JLY360, JLY361, JLY387, JLY429, JLY431, JLY433, JLY435. The *AAP1* hybrid clone was called pJL720. The *AAP1* gene was later renamed *ORC6.2*

12. The *ARS* function of the mutant sequences was analyzed in the context of *ARS1* domain B (BglII-HinfI fragment, nt 853-734) in the following CEN-based
25 URA3-containing plasmids: pJL347 (wt), pJL243 (multiple), pJL326 (A863T), pJL338 (T869A), pJL330 (T862C), and pJL316 (T867G). These plasmids were transformed into JLY106 (*MATa ura3 leu2 his3 trp1 lys2 ade2*) and its homozygous diploid counterpart JLY162. pJL243, pJL326, and pJL338 did not yield a high frequency of transformation and could not be assayed quantitatively
30 for *ARS* function. pJL347, pJL330, and pJL316 transformed cells with high efficiency and were assayed for mitotic stability [Stinchcomb, et al. *Nature* 282, 39 (1979)].

13. pJL720, the *ORC6* hybrid construct originally isolated from the YL3 library, has two BamHI sites. The 5' site created by the hybrid junction corresponds to Sau3a site at nt. 843. Excision of the segment between the two sites generated pJL721, leaving amino acids 339-435 in frame with the GAL4^{AD}.
- 5 pGAD3R (11) the parent vector for the YL3 library, contains no *ORC6* sequence. pRS425, Christianson, et al., *Gene* 110, 119 (1992), contains no components of the fusion protein.
14. All sequencing was performed with Sequenase (USB) on collapsed double-stranded templates. The protein coding segments of the *AAP1-5* hybrid clones were
 - 10 sequenced from their junction with the GAL4^{AD} to their stop codon. Two of the *ORC6* sequencing primers were used as colony hybridization probes to screen a high copy number yeast genomic library [M. Carlson and D. Botstein, *Cell* 28, 145 (1982)] for a clone of the full-length *ORC6* gene (pJL724). The full-length gene was sequenced on both strands using oligonucleotide primers positioned
 - 15 approximately 200 nt apart.
 15. S. P. Bell and B. Stillman, *Nature* 357, 128 (1992).
 16. Hodgman, *Nature* 333, 22 (1988); Walker et al., *EMBO J.* 1, 945 (1982).
 17. P. Linder, et al., *Nature* 337, 121 (1989).
 18. E. A. Nigg, *Seminars in Cell Biology* 2, 261 (1991).
 - 20 19. *ORC6* deletions were constructed by replacing nucleotides 458-1721 (pJL731) or nucleotides 458-846 (pJL733) of the Genbank sequence with the URA3 HindIII fragment oriented in the opposite direction to that of the *ORC6* sequence. Each construct was used to generate heterozygous deletions of *ORC6* in diploid strains by one-step gene replacement. *ORC6* deletion analysis was performed in
 - 25 JLY461 (*MATa/MATa ura3/ura3 leu2/leu2 his3/his3 trp1/trp1 ade2/ade2 [cir⁺]*), JLY462 (*MATa/MATa ura3/ura3 leu2/leu2 trp1/trp1 his4/his4 can1/can1*), and JLY463 (*MATa/MATa ura3/ura3 leu2/leu2 trp1/trp1 his3/HIS3*); their respective genetic backgrounds are S288c, EG123, and A364a. Disruption of JLY461, JLY462, and JLY463 by pJL731 (full deletion) created JLY481, JLY475, and
 - 30 JLY469, respectively. Disruption of JLY461, JLY462, and JLY463 by pJL733 (N-terminal deletion) created JLY485, JLY479, JLY473, respectively. These

heterozygous marked deletion strains were sporulated, and twenty tetrads of each were dissected and grown on YEPD to assess viability.

20. Pringle and Hartwell, in *The Molecular Biology of the Yeast Saccharomyces* Strathern, et al, Eds. (CSHL Press, CSH, 1981), vol. 1, pp. 97-142.
- 5 21. A point mutant (pJL766) was made by replacing the BamHI-SphI fragment of the full-length clone with a BamHI/SphI fragment generated by PCR from pJL720 using primers. One mutation changes nucleotide 1471 of the Genbank sequence from C to T and was confirmed by sequence analysis.
22. M. M. Wang and R. R. Reed, *Nature* 364, 121 (1993).
- 10 23. T. E. Wilson, et al, *Science* 252, 1296 (1991).
24. J. F. X. Diffley and J. H. Cocker, *Nature* 357, 169 (1992).
25. pJL749 contains the *GAL1* promoter (nt 146-816) driving the expression of *ORC6* (nt 443-2298) in the high-copy yeast shuttle vector RS425 [T. W. Christianson, et al., *Gene* 110, 119 (1992)].
- 15 26. The *cdc* mutant strains have been backcrossed 4-5 times against two congenic strains derived from *A364a*, RDY487 (*MAT α leu2 ura3 trp1*) and RDY488 (*MAT α leu2 ura3 trp1*). All are *ura3 leu2 trp1*. RDY510, RDY664, JLY310, and JLY179 are *MAT α* ; the rest are *MAT α* . Additional markers can be found in JLY310(*ade2*), RDY543(*his3*), and RDY619 (*pep4D::TRP1 his3 ade2*).
- 20 pJL749, pJL772, and RS425 (28) were transformed into these strains and plated on SD-LEU at 22° C. Four colony-purified isolates from each transformation were patched onto SD-LEU plates and replica-plated to SGAL-LEU plates, all at 22° C. The patches on SGAL-LEU were replica-plated to a series of pre-warmed SGAL-LEU plates at 22°, 25°, 27°, 30°, 32.5°, 35°, 37°, and 38° C. The viability of *cdc*
- 25 mutants containing pJL749 was compared to those containing pJL772 and pRS425.
27. Hartwell, *JMB* 104, 803 (1976); Hennessy, et al *G&D* 4, 2252(1990).
28. Chen, et al., *PNAS* 89, 10459 (1992); Hogan, et al, *ibid.* 89, 3098.
29. B.J. Andrews and S.W. Mason, *Science*. 261, 1543 (1993).

30 Example 4. Orc protein purification and gene cloning

Protein Purification: To obtain sufficient protein for peptide sequencing, a revised purification procedure for ORC was devised, based on the procedure reported previously (Bell and Stillman, 1992). Whole cell extract was

- prepared from 400g of frozen BJ926 cells (frozen immediately after harvesting a 300 liter logarithmically growing culture, total of 1.6 kg per 300 liters). All buffers contained 0.5 mM PMSF, 1 mM benzamidine, 2 mM pepstatin A, 0.1 mg/ml bacitracin and 2mM DTT. 400 mls of 2X buffer H/0.1^{NP-40} (100 mM
- 5 Hepes-KOH, pH 7.5, 0.2 M KCl, 2 mM EDTA, 2 mM EGTA, 10 mM Mg Acetate, and 20% glycerol) was added to the cells and after thawing the cells were broken using a bead beater (Biospec Products) until greater than 90% cell breakage was achieved (twenty 30 second pulses separated by 90 second pauses). After breakage is complete, the volume of the broken cells was measured and one twelfth
- 10 volume of a saturated (at 4°C) solution of ammonium sulfate was added and stirred for 30 minutes. This solution was then spun at 13,000 x g for 20 minutes. The resulting supernatant was transferred to 45Ti bottle assemblies (Beckman) and spun in a 45Ti rotor at 44,000 RPM for 1.5 hrs. The volume of the resulting supernatant was measured and 0.27g/ml of ammonium sulfate was added. After
- 15 stirring for 30 minutes, the precipitate was collected by spinning in the 45 Ti rotor at 40,000 RPM for 30 minutes. The resulting pellet was resuspended using a B-pestle dounce in buffer H/0.0 (50 mM Hepes-KOH, pH 7.5, 1 mM EDTA, 1 mM EGTA, 5 mM Mg Acetate, 0.02% NP-40, 10% glycerol) and dialyzed versus H/0.15M KCl (Buffer H with 0.15 M KCl added). This preparation typically
- 20 yielded 12-16 g soluble protein (determined by Bradford assay with a bovine serum albumin standard). Preparation of ORC from this extract was essentially as described (Bell and Stillman, 1992) with the following changes (column sizes used for preparation of ORC from 400g of cells are indicated in parenthesis). The S-Sepharose column was loaded at 20 mg protein per ml of resin (~ 300 ml). The
- 25 Q-Sepharose (50 ml) and sequence specific affinity column (5ml) was run as described but the dsDNA cellulose column was omitted from the preparation. Only a single glycerol gradient was performed in an SW-41 rotor spun at 41,000 RPM for 20 hrs. We estimate a yield of 130 µg of ORC complex (all subunits combined) per 400 g of yeast cells.
- 30 Protein Sequencing: Digestion of ORC subunits was performed using an "in gel" protocol described by Kawasaki and Suzuki with some modification. Briefly, purified ORC (~ 10 µg per subunit) was first separated by 10% SDS-PAGE and stained with 0.1% Coomassie Brilliant Blue G (Aldrich) for 15 min.

- After destaining (10% methanol, 10% acetic acid), the gel was soaked in water for one hour, then the protein bands were excised, transferred to a microcentrifuge tube and cut into 3-5 pieces to fit snugly into the bottom of the tube. A minimum volume of 0.1M Tris-HCl (pH=9.0) containing 0.1% SDS was added to
- 5 completely cover the gel pieces. Then 200 ng of *Achromobacter* protease I (Lysylendopeptidase: Wako) was added and incubated at 30°C for 24 hrs. After digestion the samples were centrifuged and the supernatant was passed through an Ultrafree-MC filter (Millipore, 0.22µm). The gel slices were then washed twice in 0.1% TFA for one hour and the washes were recovered and filtered as above.
- 10 All filtrates were combined and reduced to a volume suitable for injection on the HPLC using a speed-vac. The digests were separated by reverse-phase HPLC (Hewlett-Packard 1090 system) using a Vydac C18 column (2.1x 250 mm, 5µm, 300 angstroms) with an ion exchange pre-column (Brownlee GAX-013, 3.2x 15mm). The peptides were eluted from the C-18 column by increasing acetonitrile
- 15 concentration and monitored by their absorbance at 214, 280, 295, and 550 nm. Amino acid sequencing of the purified peptides was performed on an automated sequencer (Applied Biosystems model 470) with on-line HPLC (Applied Biosystems model 1020A) analysis of PTH-amino acids.

ORC SUBUNIT CLONING:

- 20 *ORC1*: To clone the gene for the largest (120 kd) subunit of ORC, the following degenerate oligonucleotide primers 1201 and 1202 were synthesized based on the sequence of the first ORC1 peptide. These oligos were used to perform PCR reactions using total yeast genomic DNA from the strain W303 a as target. A 48 base pair fragment was specifically amplified. This fragment was subcloned
- 25 and sequenced. The resulting sequence encoded the predicted peptide indicating that it was the correct amplification product. A radioactively labeled form of the PCR product was then used to probe a genomic library of yeast DNA sequences resulting in the identification of two overlapping clones. Sequencing of these clones resulted in the identification of a large open reading frame that encoded a
- 30 protein with a predicted molecular weight of 120 kd and that encoded all four of the ORC1 peptide sequences.

ORC3: To clone the gene for the 62 kd subunit of ORC, the following degenerate oligonucleotide primers 621 and 624 were synthesized based on the

sequence of the third peptide. These oligos were used to perform PCR reactions using total yeast genomic DNA from the strain W303 a as target. A 53 base pair fragment was specifically amplified. This fragment was subcloned and sequenced. The resulting sequence encoded the predicted peptide indicating that it was the correct amplification product. A radioactively labeled form of the PCR product was then used to probe a genomic library of yeast DNA sequences resulting in the identification of two overlapping clones. Sequencing of these clones resulted in the identification of a large open reading frame that encoded a protein with a predicted molecular weight of 71 kd and encoded all three of the ORC3 peptide sequences. The inconsistency of the molecular weight is presumably due to anomalous migration of this protein during SDS-PAGE.

ORC4: By comparing the sequence of the ORC4 peptides to that of the known potentially protein encoding sequences in the genbank database we found that a portion of the ORC4 coding sequence had been previously cloned in the process of cloning the adjacent gene. Using the information from the database we were able to design a perfect match oligo and use this to immediately screen a yeast library. Using this oligo as a probe of the same yeast genomic DNA library a lambda clone was isolated that contained the entire *ORC4* gene. This gene encoded a protein of predicted molecular weight 56 kd and also all of the peptides derived from the peptide sequencing of the 56 kd subunit.

ORC5: To clone the gene for the 53 kd subunit of ORC, the following degenerate oligonucleotide primers 535 and 536 were synthesized based on the sequence of the first ORC5 peptide. These oligos were used to perform PCR reactions using total yeast genomic DNA from the strain W303 a as target. A 47 base pair fragment was specifically amplified. This fragment was subcloned and sequenced. The resulting sequence encoded the predicted peptide indicating that it was the correct amplification product. A radioactively labeled form of the PCR product was then used to probe a genomic library of yeast DNA sequences resulting in the identification of a single lambda clone. Sequencing of this clones resulted in the identification of a large open reading frame that encoded a several of the peptide sequences derived from the 53 kd subunit of ORC indicating that this was the correct gene. However the sequence of the 5' end of the gene was not present in this lambda clone. Fortunately, the mutations in the same gene had also

been picked up in the same screen that resulted in the identification of the *ORC2* gene. A complementing clone to this mutation was found to overlap with the lambda clone and contain the entire 5' end of the gene. Sequencing of this complementing DNA fragment resulted in the identification of the entire sequence
5 of the *ORC5* gene.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. Although the foregoing invention has been described in some detail by way of
10 illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

15
SEQUENCE LISTING

(1) GENERAL INFORMATION:

- 20 (i) APPLICANT: COLD SPRING HARBOR LABORATORY
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
- (ii) TITLE OF INVENTION: ORC GENES, RECOMBINANT ORC PEPTIDES AND
25 METHODS OF IDENTIFYING DNA BINDING PROTEINS
- (iii) NUMBER OF SEQUENCES: 12
- (iv) CORRESPONDENCE ADDRESS:
30 (A) ADDRESSEE: FLEHR, HOHBACH, TEST, ALERITTON & HERBERT
(B) STREET: 4 Embarcadero Center, Suite 3400
(C) CITY: San Francisco
(D) STATE: California
(E) COUNTRY: USA
35 (F) ZIP: 94111-4187
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
40 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
45 (B) FILING DATE:
(C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Osman, Richard A
50 (B) REGISTRATION NUMBER: 36,677
(C) REFERENCE/DOCKET NUMBER: FP-59032-PC/RAO
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (415) 781-1989

(B) TELEFAX: (415) 398-3249
(C) TELEX: 910 277299

5 (2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 4940 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATAACATGCT CGCCCTTTTA TATTATGACA GAAAGAATAT ATATATTCAT ATATAAGATG 60
CTTCTATTTA TTAGTTTTAT CTTTAAATTG ATGATGTGTC CATAGAATTT AAGTAAGTGC 120
20 ATGGTATGGA GTGTATAATG GTTTATAATT TCCCCTAAGA TGACACAAAA AAATGTTCTC 180
CCAAAAATTT ACCAAGAAAA AAAATTAAGA ATACTACACA ATTGATGCTT GGGTTATTTT 240
25 AAATATCCGG TACATTCTAT TACAAATATG TTTGTACAAT GTAAGCCCCCT TCATAATGGT 300
CAGTATTAAG ATAAGGACTG CTATGGGGCA TTTTTTGTCT TACTGGGTAT CACAGGATAA 360
TAACTTGGCG CCAAATTAGA AAAGATATAA ACCTCAAATA TTTGAAATTC TTTGGTGACC 420
30 TGTCTCATCG TTATATCAAC AAATATTGCA CCAACGAACA CCACTACATA TGTAAGTACT 480
CTCTTCTCG ACTTATTTTT TATTAACGTT GACACGGCCA GATCGAAAAT CATAGAAAAA 540
35 CAACAACATT GAGAAGAGAT GAAGTTGCGC AAAGGGAAAG AAAACTGCAT AGGCGGCAAA 600
TTCAGCCTAA AAGTTTCCAG AAGCAGGAAC TCATTCCCTA TTGATTAATA CTCATTACAA 660
AAACCACAAT AGAGTAGATA AGATGGCAAA AACGTTGAAG GATTTACAGG GTTGGGAGAT 720
40 AATAACAAC TATGAGCAGG GAAATATAAT CGATGGAGGT CAGAAGAGAT TACGCCGAAG 780
AGGTGCAAAA ACTGAACATT ACTTAAAGAG AAGTTCTGAT GGAATTAAAC TAGGTCGTGG 840
45 TGATAGTGTA GTCATGCACA ACGAAGCCGC TGGGACTTAC TCCGTTTATA TGATCCAGGA 900
GTTGAGACTT AATACATTAA ATAATGTTGT CGAACTCTGG GCTCTCACCT ATTTACGATG 960
GTTTGAAGTC AATCCTTTAG CTCATTATAG GCAGTTTAAT CCTGACGCTA ACATTTTGAA 1020
50 TCGTCCTTTA AATTATTACA ATAACTGTT TTCTGAAACT GCAAATAAAA ATGAACTGTA 1080
TCTCACTGCA GAATTAGCCG AATTGCAGCT ATTTAACTTT ATCAGGGTTG CCAACGTAAT 1140
55 GGATGGAAGC AAATGGGAAG TATTGAAAGG AAATGTCGAT CCAGAAAGAG ACTTTACAGT 1200
TCGTTATATT TGTGAGCCGA CTGGGGAGAA ATTTGTGGAC ATTAATATTG AGGATGTCAA 1260
AGCTTACATA AAGAAAGTGG AGCCAAGGGA AGCCCAGGAA TATTTGAAAG ATTTAACACT 1320
60 TCCATCAAAG AAGAAAGAGA TCAAAAGAGG TCCTCAAAG AAAGATAAGG CTACTCAAAC 1380
GGCACAAATT TCAGACGCAG AAACAAGAGC TACAGATATA ACGGATAATG AGGACGGTAA 1440
65 TGAAGATGAA TCATCTGATT ATGAAAGTCC GTCAGATATC GACGTTAGCG AGGATATGGA 1500
CAGCGGTGAA ATATCCGCAG ATGAGCTTGA GGAAGAAGAA GACGAAGAAG AAGACGAAGA 1560

CGAAGAAGAG AAAGAAGCTA GGCATACAAA TTCACCAAGG AAAAGAGGCC GTAAGATAAA 1620
ACTAGGTAAA GATGATATTG ACGCTTCTGT ACAACCTCCC CCCAAAAAAA GAGGTCGTAA 1680
5 ACCTAAAGAT CCTAGTAAAC CGCGTCAGAT GCTATTGATA TCTTCATGCC GTGCAAATAA 1740
TACTCCTGTG ATTAGGAAAT TTACAAAAAA GAATGTTGCT AGGGCGAAAA AGAAATATAC 1800
10 CCCGTTTTCG AAAAGATTTA AATCTATAGC TGCAATACCA GATTTAACTT CATTACCTGA 1860
ATTTTACGGA AATTCTTCGG AATTGATGGC ATCAAGGTTT GAAAACAAAT TAAAAACAAC 1920
CCAAAAGCAT CAGATTGTAG AAACAATTTT TTCTAAAGTC AAAAAACAGT TGAACTCTTC 1980
15 GTATGTCAAA GAAGAAATAT TGAAGTCTGC AAATTTCCAA GATTATTTAC CGGCTAGGGA 2040
GAATGAATTC GCCTCAATTT ATTTAAGTGC ATATAGTGCC ATTGAGTCCG ACTCCGCTAC 2100
TACTATATAC GTGGCTGGTA CGCCTGGTGT AGGGAAAAC TTAACCGTAA GGAAGTCGT 2160
20 AAAGGAACTA CTATCGTCTT CTGCACAACG AGAAATACCA GACTTTCTTT ATGTGGAAAT 2220
AAATGGATTG AAAATGGTAA AACCCACAGA CTGTTACGAA ACTTTATGGA ACAAAGTGTC 2280
25 AGGAGAAAGG TTAACATGGG CAGCTTCAAT GGAGTCACTA GAGTTTTACT TTAAAGAGT 2340
TCCAAAAAAT AAGAAGAAA CCATTGTAGT CTTGTTGGAC GAATCGATG CCATGGTAAC 2400
GAAATCTCAA GATATTATGT ACAATTTTTT CAATTGGACT ACTTACGAAA ATGCCAACT 2460
30 TATTGTCATT GCAGTAGCCA ATACAATGGA CTTACCAGAA CGTCAGCTAG GCAATAAGAT 2520
TACTTCAAGA ATTGGGTTTA CCAGAATTAT GTTCACTGGG TATACGCACG AAGAGCTAAA 2580
35 AAATATCATT GATTTAAGAC TGAAGGGGTT GAACGACTCA TTTTCTATG TTGATACAAA 2640
AACTGGCAAT GCTATTTTGA TTGATGCGGC TGGAAACGAC ACTACAGTTA AGCAAACGTT 2700
GCCTGAAGAC GTGAGGAAAG TTCGCTTAAG AATGAGTGCT GATGCCATTG AAATAGCTTC 2760
40 GAGAAAAGTA GCAAGTGTTA GTGGTGATGC AAGAAGAGCA TTGAAGGTTT GTAAAAGAGC 2820
AGCTGAAATT GCTGAAAAAC ACTATATGGC TAAGCATGGT TATGGATATG ATGGAAAGAC 2880
45 GGTATTGAA GATGAAAATG AGGAGCAAAT ATACGATGAT GAAGACAAGG ATCTTATTGA 2940
AAGTAACAAA GCCAAAGACG ATAATGATGA CGATGATGAC AATGATGGGG TACAAACAGT 3000
TCACATCACG CACGTTATGA AAGCCTTAAA CGAAACTTTA AATTCTCATG TAATTACGTT 3060
50 TATGACGCGA CTTTCATTTA CAGCAAACT GTTTATTTAT GCATTATTAA ACTTGATGAA 3120
AAAGAACGGA TCTCAAGAGC AAGAACTGGG CGATATTGTC GATGAAATCA AGTTACTTAT 3180
55 TGAAGTAAAT GGCAGTAATA AGTTTGTCAT GGAGATAGCC AAAACATTGT TCCAACAGGG 3240
AAGTGATAAT ATTTCTGAAC AATTGAGAAT TATATCATGG GATTTCGTTT TCAATCAGTT 3300
ACTTGACGCG GGAATATTGT TTAAACAAAC TATGAAGAAC GATAGAATAT GTTGTGTCAA 3360
60 GCTAAATATA TCAGTAGAAG AAGCCAAAAG AGCCATGAAT GAGGATGAGA CATTGAGAAA 3420
TTTATAGATT CGGTTTTTAT TATTCATGAC CTAGCATACA CATACATATA CCTACATAGT 3480
65 AGCGCATTTA TCCAAAACAT ACGATATTGT GGATGTACAT ACCTTCTATA TCTCCTTAAA 3540
GCTATTGTGT AGCTTGATTT AAAATATGCT AACGCCAACT CTCACATGGT AGCAGGCGGG 3600

TATAGTTGTT TTCATGTATT AACGCCCCGC GATGGTGCCT TAGATGAGGG CGACGAGGAG 3660
 GGCTTCCTGA TATTATGGCT CTTTCTATCC TGACTTTTGT TATGATGTCG ATGTTGCTGG 3720
 5 CCACCTAGGT GCTTATATAT CAAAAGAGGA TCGCCGATTT CATTGATTTT TGGGATGGTT 3780
 AATGTCAAAT TAAAGATCTT TGCCAGTGCA ATTTTGAAAA TTTTGTGAAT GTTTATAGAT 3840
 10 TTGGCAGTAG AGCAGAATAT AAGAGGAGCA TTCATGACCT GTGCATACTT CATACTCGTT 3900
 CTCGAGATTT GTTCCTGATA TTCCGGGTCT AAGTCTATTA GTAAATCGTA CTTTGTGCCC 3960
 ACCAAAATAG GAATTGCCGA ATCATTTAGC CCGTACGCCT GCCTATACCA CTCCTTTATT 4020
 15 GAACTCAACG TCTCTGGACG TGTCAGGTCA AACAGAAATA TGATCACTGA AGACCCTACC 4080
 GTCGCAATTG GGAGCATGTT GATGAATTCT CTTTGTCCGC CTAAATCCAT TATAGAAAAT 4140
 20 ATAATATCCG TGGAGCGTAT GCTTACTTTT CTTTTCAAAA AGTTCACTCC CAGCGTCTGT 4200
 GTGTATTCCT TATCGTATAT GTTCTGTACG TACTTCACCA TCAGCGATGT TTTCCCTACT 4260
 TGTGCATCCC CTACTAATCC AACCTGAACT TCAACCTGAT TTCGTACCGC AGGTATAGAA 4320
 25 TTGTTTGCTC CCGTGCTTGG TGTAGCCATC TTAGCTTAAC TCAATTTAAT TTCTACAGCA 4380
 AAATCCAAAC GTAATATCTA TATTTTCTC GAAAACTGA GGACAAGAGC CAATCAATCA 4440
 30 TCTATAATCC AATTTATATT ATTTTTCCTC TTTCTGGTTC TTTTCTTCCT TTTCTGTTT 4500
 ACCTTTTTTG CTTTTTCATA AAATAATTTT TCTAGATTTG AAGACAGCAT TTTGTACAT 4560
 CCATACACCA TACACCATAC ACCATAGCAC CAGTACACTA TATTTTTATG AATTTTACTA 4620
 35 AGAATTATTC CTGCAGGAGC TCCACTGAAA AAAAAAGAGC AGCATGGATG TCATGTCGGT 4680
 AGAGTGCTAC TGAGTAAATG GGAGGACGCG GTAGATCCAG TGTGGAATCA AGGTGGTGCC 4740
 40 GGTGTGAAGC CGCCTCGGCC GGCTGGACTC TCCAGGCCGG AGTGATGATT GCCACGCTGA 4800
 AGCTAACACA GTTTCACAAT ACCAGTGTCC TCATTAGTGA GTTCCAATGT ATAGTTAGTA 4860
 GTGGTATTTT GATATATGTG AGTGGTAGCA GATTGAACT TAGTTAGTTG TATTCGCCTT 4920
 45 TGAGGAAACC AAGCCAAAAA 4940

(2) INFORMATION FOR SEQ ID NO:2:

- 50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 914 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 55 (ii) MOLECULE TYPE: peptide
 60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
 Met Ala Lys Thr Leu Lys Asp Leu Gln Gly Trp Glu Ile Ile Thr Thr
 1 5 10 15
 65 Asp Glu Gln Gly Asn Ile Ile Asp Gly Gly Gln Lys Arg Leu Arg Arg
 20 25 30
 Arg Gly Ala Lys Thr Glu His Tyr Leu Lys Arg Ser Ser Asp Gly Ile
 35 40 45

	Lys	Leu	Gly	Arg	Gly	Asp	Ser	Val	Val	Met	His	Asn	Glu	Ala	Ala	Gly	
	50						55					60					
5	Thr	Tyr	Ser	Val	Tyr	Met	Ile	Gln	Glu	Leu	Arg	Leu	Asn	Thr	Leu	Asn	
	65					70					75					80	
	Asn	Val	Val	Glu	Leu	Trp	Ala	Leu	Thr	Tyr	Leu	Arg	Trp	Phe	Glu	Val	
				85						90					95		
10	Asn	Pro	Leu	Ala	His	Tyr	Arg	Gln	Phe	Asn	Pro	Asp	Ala	Asn	Ile	Leu	
				100					105					110			
	Asn	Arg	Pro	Leu	Asn	Tyr	Tyr	Asn	Lys	Leu	Phe	Ser	Glu	Thr	Ala	Asn	
			115					120					125				
15	Lys	Asn	Glu	Leu	Tyr	Leu	Thr	Ala	Glu	Leu	Ala	Glu	Leu	Gln	Leu	Phe	
	130						135					140					
20	Asn	Phe	Ile	Arg	Val	Ala	Asn	Val	Met	Asp	Gly	Ser	Lys	Trp	Glu	Val	
	145					150					155					160	
	Leu	Lys	Gly	Asn	Val	Asp	Pro	Glu	Arg	Asp	Phe	Thr	Val	Arg	Tyr	Ile	
				165						170					175		
25	Cys	Glu	Pro	Thr	Gly	Glu	Lys	Phe	Val	Asp	Ile	Asn	Ile	Glu	Asp	Val	
				180					185					190			
	Lys	Ala	Tyr	Ile	Lys	Lys	Val	Glu	Pro	Arg	Glu	Ala	Gln	Glu	Tyr	Leu	
			195					200					205				
30	Lys	Asp	Leu	Thr	Leu	Pro	Ser	Lys	Lys	Lys	Glu	Ile	Lys	Arg	Gly	Pro	
	210						215					220					
35	Gln	Lys	Lys	Asp	Lys	Ala	Thr	Gln	Thr	Ala	Gln	Ile	Ser	Asp	Ala	Glu	
	225					230					235					240	
	Thr	Arg	Ala	Thr	Asp	Ile	Thr	Asp	Asn	Glu	Asp	Gly	Asn	Glu	Asp	Glu	
				245						250					255		
40	Ser	Ser	Asp	Tyr	Glu	Ser	Pro	Ser	Asp	Ile	Asp	Val	Ser	Glu	Asp	Met	
				260					265					270			
	Asp	Ser	Gly	Glu	Ile	Ser	Ala	Asp	Glu	Leu	Glu	Glu	Glu	Glu	Asp	Glu	
			275					280					285				
45	Glu	Glu	Asp	Glu	Asp	Glu	Glu	Glu	Lys	Glu	Ala	Arg	His	Thr	Asn	Ser	
	290						295					300					
50	Pro	Arg	Lys	Arg	Gly	Arg	Lys	Ile	Lys	Leu	Gly	Lys	Asp	Asp	Ile	Asp	
	305					310					315					320	
	Ala	Ser	Val	Gln	Pro	Pro	Pro	Lys	Lys	Arg	Gly	Arg	Lys	Pro	Lys	Asp	
				325						330					335		
55	Pro	Ser	Lys	Pro	Arg	Gln	Met	Leu	Leu	Ile	Ser	Ser	Cys	Arg	Ala	Asn	
				340					345					350			
	Asn	Thr	Pro	Val	Ile	Arg	Lys	Phe	Thr	Lys	Lys	Asn	Val	Ala	Arg	Ala	
			355					360					365				
60	Lys	Lys	Lys	Tyr	Thr	Pro	Phe	Ser	Lys	Arg	Phe	Lys	Ser	Ile	Ala	Ala	
	370						375					380					
65	Ile	Pro	Asp	Leu	Thr	Ser	Leu	Pro	Glu	Phe	Tyr	Gly	Asn	Ser	Ser	Glu	
	385					390					395					400	
	Leu	Met	Ala	Ser	Arg	Phe	Glu	Asn	Lys	Leu	Lys	Thr	Thr	Gln	Lys	His	
				405						410					415		

Gln Ile Val Glu Thr Ile Phe Ser Lys Val Lys Lys Gln Leu Asn Ser
 420 425 430
 5 Ser Tyr Val Lys Glu Glu Ile Leu Lys Ser Ala Asn Phe Gln Asp Tyr
 435 440 445
 Leu Pro Ala Arg Glu Asn Glu Phe Ala Ser Ile Tyr Leu Ser Ala Tyr
 450 455 460
 10 Ser Ala Ile Glu Ser Asp Ser Ala Thr Thr Ile Tyr Val Ala Gly Thr
 465 470 475 480
 Pro Gly Val Gly Lys Thr Leu Thr Val Arg Glu Val Val Lys Glu Leu
 485 490 495
 15 Leu Ser Ser Ser Ala Gln Arg Glu Ile Pro Asp Phe Leu Tyr Val Glu
 500 505 510
 Ile Asn Gly Leu Lys Met Val Lys Pro Thr Asp Cys Tyr Glu Thr Leu
 515 520 525
 20 Trp Asn Lys Val Ser Gly Glu Arg Leu Thr Trp Ala Ala Ser Met Glu
 530 535 540
 25 Ser Leu Glu Phe Tyr Phe Lys Arg Val Pro Lys Asn Lys Lys Lys Thr
 545 550 555 560
 Ile Val Val Leu Leu Asp Glu Leu Asp Ala Met Val Thr Lys Ser Gln
 565 570 575
 30 Asp Ile Met Tyr Asn Phe Phe Asn Trp Thr Thr Tyr Glu Asn Ala Lys
 580 585 590
 35 Leu Ile Val Ile Ala Val Ala Asn Thr Met Asp Leu Pro Glu Arg Gln
 595 600 605
 Leu Gly Asn Lys Ile Thr Ser Arg Ile Gly Phe Thr Arg Ile Met Phe
 610 615 620
 40 Thr Gly Tyr Thr His Glu Glu Leu Lys Asn Ile Ile Asp Leu Arg Leu
 625 630 635 640
 Lys Gly Leu Asn Asp Ser Phe Phe Tyr Val Asp Thr Lys Thr Gly Asn
 645 650 655
 45 Ala Ile Leu Ile Asp Ala Ala Gly Asn Asp Thr Thr Val Lys Gln Thr
 660 665 670
 50 Leu Pro Glu Asp Val Arg Lys Val Arg Leu Arg Met Ser Ala Asp Ala
 675 680 685
 Ile Glu Ile Ala Ser Arg Lys Val Ala Ser Val Ser Gly Asp Ala Arg
 690 695 700
 55 Arg Ala Leu Lys Val Cys Lys Arg Ala Ala Glu Ile Ala Glu Lys His
 705 710 715 720
 Tyr Met Ala Lys His Gly Tyr Gly Tyr Asp Gly Lys Thr Val Ile Glu
 725 730 735
 60 Asp Glu Asn Glu Glu Gln Ile Tyr Asp Asp Glu Asp Lys Asp Leu Ile
 740 745 750
 65 Glu Ser Asn Lys Ala Lys Asp Asp Asn Asp Asp Asp Asp Asp Asn Asp
 755 760 765
 Gly Val Gln Thr Val His Ile Thr His Val Met Lys Ala Leu Asn Glu
 770 775 780

Thr Leu Asn Ser His Val Ile Thr Phe Met Thr Arg Leu Ser Phe Thr
 785 790 795 800
 5 Ala Lys Leu Phe Ile Tyr Ala Leu Leu Asn Leu Met Lys Lys Asn Gly
 805 810 815
 Ser Gln Glu Gln Glu Leu Gly Asp Ile Val Asp Glu Ile Lys Leu Leu
 820 825 830
 10 Ile Glu Val Asn Gly Ser Asn Lys Phe Val Met Glu Ile Ala Lys Thr
 835 840 845
 Leu Phe Gln Gln Gly Ser Asp Asn Ile Ser Glu Gln Leu Arg Ile Ile
 850 855 860
 15 Ser Trp Asp Phe Val Leu Asn Gln Leu Leu Asp Ala Gly Ile Leu Phe
 865 870 875 880
 20 Lys Gln Thr Met Lys Asn Asp Arg Ile Cys Cys Val Lys Leu Asn Ile
 885 890 895
 Ser Val Glu Glu Ala Lys Arg Ala Met Asn Glu Asp Glu Thr Leu Arg
 900 905 910
 25 Asn Leu

(2) INFORMATION FOR SEQ ID NO:3:

- 30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2809 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 35 (ii) MOLECULE TYPE: cDNA
 (ix) FEATURE:
 (A) NAME/KEY: CDS
 40 (B) LOCATION: 807..2666
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

45 GAGCTCAACA CCACCATTGA GAACGTAGAA TTTCAATTTT TAAGCTGATT CTCTTTCTGC 60
 ATGAACTCTC CTAGCAATGT GAAACTTCTC TTAAGGGAAA TTTTCGCCTT TTTGAATGGG 120
 CATACTTGGC CAAAATTCA GGATTGAATA TATATAATCG GAAGTTGTAT GGATAAAAAAT 180
 50 TTATATCAAG AGTCTGTTTC TTAATTGGAT TTGCTGTGAT CTAGTATTGA GATGACTATA 240
 AACCGGCCAG GAAATTAGTC TTTTCGAAGC TGGTTTTGGT TTCGCAAGAG TCTTTTTGAC 300
 AGCTTTTTGG CCTCAATTTG TATTCCCTTA ATACGCTTCT TCAACTCTGT CTTAGAGACC 360
 55 ATTTCTCCAG TGGCCTCATC TAGGTGTAAA CTAGCAATAG CGTCACTAGC TGCCGTGACA 420
 TTAAGTTGCT GTGGCACCTT TATATGTAAT ATGAACCATC TTTCAATGGA TCATAAGAAT 480
 60 AAGTGTCGTA AAAGGCCAAA TATCCATGCA TAAATATCGA CTTATTCCGG TAAATGTGAT 540
 ATGGATCAGC TAGTACCAAT TTCTAGTCTA GCAAATCGG GAAAATTTTT CAGAACACCC 600
 ACTCACCGCA TCATTGAGGT GGAAATGACA ATAGTAAGCA GAATTGTTAT TCTTCACAAT 660
 65 GTGTAAAAGT TATAAAGAAA TAGGAACCAC CTTTAAATTA AGACAAAGTA GAATATATTA 720
 GCTGAAATTG TATTTGATAA TTGATCATTG ATCTTATTG CTATATCTTT AAAACAAGTT 780

	TTTGTAGTAC TGC GAATTGC CATAAC ATG CTA AAT GGG GAA GAC TTT GTA GAG	833
	Met Leu Asn Gly Glu Asp Phe Val Glu	
	1 5	
5	CAT AAT GAT ATC CTA TCG TCT CCG GCA AAA AGC AGG AAT GTA ACC CCA	881
	His Asn Asp Ile Leu Ser Ser Pro Ala Lys Ser Arg Asn Val Thr Pro	
	10 15 20 25	
10	AAA AGG GTT GAC CCA CAT GGA GAA AGA CAA CTG AGA AGA ATT CAT TCA	929
	Lys Arg Val Asp Pro His Gly Glu Arg Gln Leu Arg Arg Ile His Ser	
	30 35 40	
15	TCA AAG AAG AAT TTG TTG GAA AGA ATC TCG CTT GTA GGC AAC GAA AGG	977
	Ser Lys Lys Asn Leu Leu Glu Arg Ile Ser Leu Val Gly Asn Glu Arg	
	45 50 55	
20	AAA AAT ACA TCT CCA GAT CCG GCA CTC AAA CCT AAA ACG CCA AGT AAA	1025
	Lys Asn Thr Ser Pro Asp Pro Ala Leu Lys Pro Lys Thr Pro Ser Lys	
	60 65 70	
25	GCT CCC CGT AAA CGT GGA AGA CCA AGA AAG ATA CAG GAA GAA TTA ACT	1073
	Ala Pro Arg Lys Arg Gly Arg Pro Arg Lys Ile Gln Glu Glu Leu Thr	
	75 80 85	
30	GAT AGG ATC AAG AAG GAT GAG AAA GAT ACA ATT TCC TCT AAG AAA AAG	1121
	Asp Arg Ile Lys Lys Asp Glu Lys Asp Thr Ile Ser Ser Lys Lys Lys	
	90 95 100 105	
35	AGG AAA TTG GAC AAA GAT ACA TCA GGT AAT GTC AAT GAG GAA AGC AAG	1169
	Arg Lys Leu Asp Lys Asp Thr Ser Gly Asn Val Asn Glu Glu Ser Lys	
	110 115 120	
40	ACT TCT AAC AAC AAG CAG GTG ATG GAA AAG ACG GGG ATA AAA GAG AAA	1217
	Thr Ser Asn Asn Lys Gln Val Met Glu Lys Thr Gly Ile Lys Glu Lys	
	125 130 135	
45	AGA GAA CGC GAA AAA ATA CAG GTA GCG ACC ACA ACA TAT GAA GAT AAT	1265
	Arg Glu Arg Glu Lys Ile Gln Val Ala Thr Thr Thr Tyr Glu Asp Asn	
	140 145 150	
50	GTG ACT CCA CAA ACT GAT GAT AAT TTT GTA TCA AAT TCA CCC GAG CCA	1313
	Val Thr Pro Gln Thr Asp Asp Asn Phe Val Ser Asn Ser Pro Glu Pro	
	155 160 165	
55	CCA GAA CCT GCA ACA CCA TCT AAG AAG TCT TTA ACC ACT AAT CAT GAT	1361
	Pro Glu Pro Ala Thr Pro Ser Lys Lys Ser Leu Thr Thr Asn His Asp	
	170 175 180 185	
60	TTT ACT TCG CCC CTA AAG CAA ATT ATA ATG AAT AAT TTA AAA GAA TAT	1409
	Phe Thr Ser Pro Leu Lys Gln Ile Ile Met Asn Asn Leu Lys Glu Tyr	
	190 195 200	
65	AAA GAC TCA ACC TCC CCA GGT AAA TTA ACC TTG AGT AGA AAT TTT ACT	1457
	Lys Asp Ser Thr Ser Pro Gly Lys Leu Thr Leu Ser Arg Asn Phe Thr	
	205 210 215	
70	CCA ACC CCT GTA CCG AAA AAT AAA AAG CTC TAC CAA ACT TCG GAA ACC	1505
	Pro Thr Pro Val Pro Lys Asn Lys Lys Leu Tyr Gln Thr Ser Glu Thr	
	220 225 230	
75	AAG TCA GCA AGC TCG TTT TTG GAT ACT TTT GAA GGA TAT TTC GAC CAA	1553
	Lys Ser Ala Ser Ser Phe Leu Asp Thr Phe Glu Gly Tyr Phe Asp Gln	
	235 240 245	
80	AGA AAA ATT GTC AGA ACT AAT GCG AAG TCA AGG CAC ACC ATG TCA ATG	1601
	Arg Lys Ile Val Arg Thr Asn Ala Lys Ser Arg His Thr Met Ser Met	
	250 255 260 265	

	GCA CCT GAC GTT ACC AGA GAA GAG TTT TCC CTA GTA TCA AAC TTT TTC	1649
	Ala Pro Asp Val Thr Arg Glu Glu Phe Ser Leu Val Ser Asn Phe Phe	
	270 275 280	
5	AAC GAA AAT TTT CAA AAA CGT CCC AGG CAA AAG TTA TTT GAA ATT CAG	1697
	Asn Glu Asn Phe Gln Lys Arg Pro Arg Gln Lys Leu Phe Glu Ile Gln	
	285 290 295	
10	AAA AAA ATG TTT CCC CAG TAT TGG TTT GAA TTG ACT CAA GGA TTC TCC	1745
	Lys Lys Met Phe Pro Gln Tyr Trp Phe Glu Leu Thr Gln Gly Phe Ser	
	300 305 310	
15	TTA TTA TTT TAT GGT GTA GGT TCG AAA CGT AAT TTT TTG GAA GAG TTT	1793
	Leu Leu Phe Tyr Gly Val Gly Ser Lys Arg Asn Phe Leu Glu Glu Phe	
	315 320 325	
20	GCC ATT GAC TAC TTG TCT CCG AAA ATC GCG TAC TCG CAA CTG GCT TAT	1841
	Ala Ile Asp Tyr Leu Ser Pro Lys Ile Ala Ser Gln Leu Ala Tyr	
	330 335 340 345	
	GAG AAT GAA TTA CAA CAA AAC AAA CCT GTA AAT TCC ATC CCA TGC CTT	1889
	Glu Asn Glu Leu Gln Gln Asn Lys Pro Val Asn Ser Ile Pro Cys Leu	
	350 355 360	
25	ATT TTA AAT GGT TAC AAC CCT AGC TGT AAC TAT CGT GAC GTC TTC AAA	1937
	Ile Leu Asn Gly Tyr Asn Pro Ser Cys Asn Tyr Arg Asp Val Phe Lys	
	365 370 375	
30	GAG ATT ACC GAT CTT TTG GTC CCC GCT GAG TTG ACA AGA AGC GAA ACT	1985
	Glu Ile Thr Asp Leu Leu Val Pro Ala Glu Leu Thr Arg Ser Glu Thr	
	380 385 390	
35	AAG TAC TGG GGC AAT CAT GTG ATT TTG CAG ATC CAA AAG ATG ATT GAT	2033
	Lys Tyr Trp Gly Asn His Val Ile Leu Gln Ile Gln Lys Met Ile Asp	
	395 400 405	
40	TTC TAC AAA AAT CAA CCT TTA GAT ATC AAA TTA ATA CTT GTA GTG CAT	2081
	Phe Tyr Lys Asn Gln Pro Leu Asp Ile Lys Leu Ile Leu Val Val His	
	410 415 420 425	
	AAT CTG GAT GGT CCT AGC ATA AGG AAA AAC ACT TTT CAG ACG ATG CTA	2129
	Asn Leu Asp Gly Pro Ser Ile Arg Lys Asn Thr Phe Gln Thr Met Leu	
	430 435 440	
45	AGC TTC CTC TCC GTC ATC AGA CAA ATC GCC ATA GTC GCC TCT ACA GAC	2177
	Ser Phe Leu Ser Val Ile Arg Gln Ile Ala Ile Val Ala Ser Thr Asp	
	445 450 455	
50	CAC ATT TAC GCT CCG CTC CTC TGG GAC AAC ATG AAG GCC CAA AAC TAC	2225
	His Ile Tyr Ala Pro Leu Leu Trp Asp Asn Met Lys Ala Gln Asn Tyr	
	460 465 470	
55	AAC TTT GTC TTT CAT GAT ATT TCG AAT TTT GAA CCG TCG ACA GTC GAG	2273
	Asn Phe Val Phe His Asp Ile Ser Asn Phe Glu Pro Ser Thr Val Glu	
	475 480 485	
60	TCT ACG TTC CAA GAT GTG ATG AAG ATG GGT AAA AGC GAT ACC AGC AGT	2321
	Ser Thr Phe Gln Asp Val Met Lys Met Gly Lys Ser Asp Thr Ser Ser	
	490 495 500 505	
	GGT GCT GAA GGT GCG AAA TAC GTC TTA CAA TCA CTT ACT GTG AAC TCC	2369
	Gly Ala Glu Gly Ala Lys Tyr Val Leu Gln Ser Leu Thr Val Asn Ser	
	510 515 520	
65	AAG AAG ATG TAT AAG TTG CTT ATT GAA ACA CAA ATG CAG AAT ATG GGG	2417
	Lys Lys Met Tyr Lys Leu Leu Ile Glu Thr Gln Met Gln Asn Met Gly	
	525 530 535	

AAT CTA TCC GCT AAC ACA GGT CCT AAG CGT GGT ACT CAA AGA ACT GGA 2465
 Asn Leu Ser Ala Asn Thr Gly Pro Lys Arg Gly Thr Gln Arg Thr Gly
 540 545 550

5 GTA GAA CTT AAA CTT TTC AAC CAT CTC TGT GCC GCT GAT TTT ATT GCT 2513
 Val Glu Leu Lys Leu Phe Asn His Leu Cys Ala Ala Asp Phe Ile Ala
 555 560 565

10 TCT AAT GAG ATA GCT CTA AGG TCG ATG CTT AGA GAA TTC ATA GAA CAT 2561
 Ser Asn Glu Ile Ala Leu Arg Ser Met Leu Arg Glu Phe Ile Glu His
 570 575 580 585

15 AAA ATG GCC AAC ATA ACT AAG AAC AAT TCT GGA ATG GAA ATT ATT TGG 2609
 Lys Met Ala Asn Ile Thr Lys Asn Asn Ser Gly Met Glu Ile Ile Trp
 590 595 600

20 GTA CCC TAC ACG TAT GCG GAA CTT GAA AAA CTT CTG AAA ACC GTT TTA 2657
 Val Pro Tyr Thr Tyr Ala Glu Leu Glu Lys Leu Leu Lys Thr Val Leu
 605 610 615

AAT ACT CTA TAAATGTATA CATATCACGA ACAATTGTAA TAGTACTAGG 2706
 Asn Thr Leu
 620

25 CTTGCTAGCT TTGCTTTCCC ATAACCAACA ATACTTAGTG ATGTATCTTA AAACGACTAA 2766
 AAAACTTCTC ATATAACCCT ACTGAAAAAC GTCTGATGAG CTC 2809

30 (2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 620 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

40 Met Leu Asn Gly Glu Asp Phe Val Glu His Asn Asp Ile Leu Ser Ser
 1 5 10 15

45 Pro Ala Lys Ser Arg Asn Val Thr Pro Lys Arg Val Asp Pro His Gly
 20 25 30

Glu Arg Gln Leu Arg Arg Ile His Ser Ser Lys Lys Asn Leu Leu Glu
 35 40 45

50 Arg Ile Ser Leu Val Gly Asn Glu Arg Lys Asn Thr Ser Pro Asp Pro
 50 55 60

Ala Leu Lys Pro Lys Thr Pro Ser Lys Ala Pro Arg Lys Arg Gly Arg
 65 70 75 80

55 Pro Arg Lys Ile Gln Glu Glu Leu Thr Asp Arg Ile Lys Lys Asp Glu
 85 90 95

60 Lys Asp Thr Ile Ser Ser Lys Lys Lys Arg Lys Leu Asp Lys Asp Thr
 100 105 110

Ser Gly Asn Val Asn Glu Glu Ser Lys Thr Ser Asn Asn Lys Gln Val
 115 120 125

65 Met Glu Lys Thr Gly Ile Lys Glu Lys Arg Glu Arg Glu Lys Ile Gln
 130 135 140

	Val	Ala	Thr	Thr	Thr	Tyr	Glu	Asp	Asn	Val	Thr	Pro	Gln	Thr	Asp	Asp	145	150	155	160
5	Asn	Phe	Val	Ser	Asn	Ser	Pro	Glu	Pro	Pro	Glu	Pro	Ala	Thr	Pro	Ser	165	170	175	
	Lys	Lys	Ser	Leu	Thr	Thr	Asn	His	Asp	Phe	Thr	Ser	Pro	Leu	Lys	Gln	180	185	190	
10	Ile	Ile	Met	Asn	Asn	Leu	Lys	Glu	Tyr	Lys	Asp	Ser	Thr	Ser	Pro	Gly	195	200	205	
	Lys	Leu	Thr	Leu	Ser	Arg	Asn	Phe	Thr	Pro	Thr	Pro	Val	Pro	Lys	Asn	210	215	220	
15	Lys	Lys	Leu	Tyr	Gln	Thr	Ser	Glu	Thr	Lys	Ser	Ala	Ser	Ser	Phe	Leu	225	230	235	240
	Asp	Thr	Phe	Glu	Gly	Tyr	Phe	Asp	Gln	Arg	Lys	Ile	Val	Arg	Thr	Asn	245	250	255	
20	Ala	Lys	Ser	Arg	His	Thr	Met	Ser	Met	Ala	Pro	Asp	Val	Thr	Arg	Glu	260	265	270	
25	Glu	Phe	Ser	Leu	Val	Ser	Asn	Phe	Phe	Asn	Glu	Asn	Phe	Gln	Lys	Arg	275	280	285	
	Pro	Arg	Gln	Lys	Leu	Phe	Glu	Ile	Gln	Lys	Lys	Met	Phe	Pro	Gln	Tyr	290	295	300	
30	Trp	Phe	Glu	Leu	Thr	Gln	Gly	Phe	Ser	Leu	Leu	Phe	Tyr	Gly	Val	Gly	305	310	315	320
	Ser	Lys	Arg	Asn	Phe	Leu	Glu	Glu	Phe	Ala	Ile	Asp	Tyr	Leu	Ser	Pro	325	330	335	
35	Lys	Ile	Ala	Tyr	Ser	Gln	Leu	Ala	Tyr	Glu	Asn	Glu	Leu	Gln	Gln	Asn	340	345	350	
40	Lys	Pro	Val	Asn	Ser	Ile	Pro	Cys	Leu	Ile	Leu	Asn	Gly	Tyr	Asn	Pro	355	360	365	
	Ser	Cys	Asn	Tyr	Arg	Asp	Val	Phe	Lys	Glu	Ile	Thr	Asp	Leu	Leu	Val	370	375	380	
45	Pro	Ala	Glu	Leu	Thr	Arg	Ser	Glu	Thr	Lys	Tyr	Trp	Gly	Asn	His	Val	385	390	395	400
	Ile	Leu	Gln	Ile	Gln	Lys	Met	Ile	Asp	Phe	Tyr	Lys	Asn	Gln	Pro	Leu	405	410	415	
50	Asp	Ile	Lys	Leu	Ile	Leu	Val	Val	His	Asn	Leu	Asp	Gly	Pro	Ser	Ile	420	425	430	
55	Arg	Lys	Asn	Thr	Phe	Gln	Thr	Met	Leu	Ser	Phe	Leu	Ser	Val	Ile	Arg	435	440	445	
	Gln	Ile	Ala	Ile	Val	Ala	Ser	Thr	Asp	His	Ile	Tyr	Ala	Pro	Leu	Leu	450	455	460	
60	Trp	Asp	Asn	Met	Lys	Ala	Gln	Asn	Tyr	Asn	Phe	Val	Phe	His	Asp	Ile	465	470	475	480
	Ser	Asn	Phe	Glu	Pro	Ser	Thr	Val	Glu	Ser	Thr	Phe	Gln	Asp	Val	Met	485	490	495	
65	Lys	Met	Gly	Lys	Ser	Asp	Thr	Ser	Ser	Gly	Ala	Glu	Gly	Ala	Lys	Tyr	500	505	510	

Val Leu Gln Ser Leu Thr Val Asn Ser Lys Lys Met Tyr Lys Leu Leu
 515 520 525

5 Ile Glu Thr Gln Met Gln Asn Met Gly Asn Leu Ser Ala Asn Thr Gly
 530 535 540

Pro Lys Arg Gly Thr Gln Arg Thr Gly Val Glu Leu Lys Leu Phe Asn
 545 550 555 560

10 His Leu Cys Ala Ala Asp Phe Ile Ala Ser Asn Glu Ile Ala Leu Arg
 565 570 575

Ser Met Leu Arg Glu Phe Ile Glu His Lys Met Ala Asn Ile Thr Lys
 580 585 590

15 Asn Asn Ser Gly Met Glu Ile Ile Trp Val Pro Tyr Thr Tyr Ala Glu
 595 600 605

20 Leu Glu Lys Leu Leu Lys Thr Val Leu Asn Thr Leu
 610 615 620

(2) INFORMATION FOR SEQ ID NO:5:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2759 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 30 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

35 TCTGAAATAA AAAGTACAAA AAAGAAAACA ATATACCAGA TATGAACCCT TTTAGTGAGA 60
 TTCCAGCATG TCTTTGCGCA GATCCAAATC TTTCTTTGTC TTGAAATTTA TTCAGTAAAT 120
 TAAAAGTCAG TTCTTTAGTA GCATTCACTT TCTTGGTAAG TCTTTTCTT GTTTTTGAAA 180
 40 AAGAGTTCCT GAAGTTTGTC TACTGTGAAT ATACTTTGCA CATTGTGTTA ATTTTAAAC 240
 ACGCTATAAT TTGTGTCATA AAGAATTTTT TGTAGAATAG CTTTTTTTTT AATAGGAAAA 300
 45 AAAAATAAAA AAAGGTGGAA AAGACAATCT TTTCCAGAAA CTGAAACTA TACTGGAGAT 360
 GAAGGGTTGT CGTTGGTTGC GTTACGAGAC AGGCTTGACA ATTCACAAG AGTAATGTTT 420
 CATTACCTGC TGTTTTATTA TCTTTATATT TAGTAAGACC AGCAGAAACG CTACACGTGA 480
 50 TGATAATGGA ACTAAGCATT CTGTTAGATG GTAAGAATTT TTTTACCTT CCATTACCAC 540
 TAACGCCTTT TTTAGTGTCT TTTTGATATT TACTGACGTA TTTTCCGCA CCGTAATTTG 600
 55 AAGAAAAAGA AAAGTGACAA AAGATGGCAT TGTTTACATA CAGAGTCGTA GTATCACAAG 660
 AGTAGTCCAA CAGGATGAGC GACCTTAACC AATCCAAAAA GATGAACGTC AGCGAGTTTG 720
 CTGACGCCCA AAGGAGCCAC TATACAGTAT ACCCCAGTTT GCCTCAAAGT AACAAAAATG 780
 60 ATAAACACAT TCCCTTTGTC AAACCTCTAT CAGGCAAAGA ATCGGAAGTG AACGTGGAAA 840
 AAAGATGGGA ATTGTATCAT CAGTTACATT CCCACTTTCA TGATCAAGTA GATCATATTA 900
 65 TCGATAATAT TGAAGCAGAC TTGAAAGCAG AGATTTCAGA CCTTTTATAT AGTGAAACTA 960
 CTCAGAAAAG GCGATGCTTT AACACTATTT TCCTATTAGG TTCAGATAGT ACGACAAAAA 1020

TTGAACTTAA AGACGAATCT TCTCGCTACA ACGTTTTGAT TGAATTGACT CCGAAAGAAT 1080
 CTCCGAATGT AAGAATGATG CTTCGTAGGT CTATGTACAA ACTTTACAGC GCAGCTGATG 1140
 5 CAGAAGAACA TCCAACATC AAGTATGAAG ACATTAACGA TGAAGATGGC GATTTTACCG 1200
 AGCAAAACAA TGATGTATCA TACGATCTGT CACTTGTGGA AAACCTTCAAA AGGCTTTTTTG 1260
 10 GAAAAGACTT AGCAATGGTA TTTAATTTTA AAGATGTAGA TTCTATTAAC TTCAACACAT 1320
 TGGATAACTT CATAATTCTA TTGAAAAGTG CCTTCAAGTA TGACCATGTT AAAATAAGTT 1380
 TAATCTTTAA TATTAATACA AACTTGTCOA ATATTGAGAA AAATTTGAGA CAATCAACCA 1440
 15 TACGACTTCT GAAGAGAAAT TATCATAAAC TAGACGTGTC GAGTAATAAA GGATTTAAGT 1500
 ACGGAAACCA AATCTTTTCAA AGCTTTTTTGG ATACGGTTGA TGGCAAACCA AATCTTTTCAG 1560
 ATCGTTTTGT GGAATTCATT CTCAGCAAGA TGGCAAATAA TACTAATCAC AACTTACAAT 1620
 20 TATTGACGAA GATGCTGGAT TATTCGTTGA TGTCGTACTT TTTCCAGAAT GCCTTTTCAG 1680
 TATTCATTGA CCCTGTAAAT GTTGATTTTT TGAACGACGA CTACTTAAAA ATACTGAGCA 1740
 25 GATGTCCTAC ATTCATGTTT TTTGTGGAAG GTCTTATAAA GCAGCATGCT CCTGCTGACG 1800
 AAATTCCTTC ATTATTGACA AACAAAAACA GAGGCCTAGA AGAGTTTTTT GTTGAGTTTT 1860
 TGGAAGAGA GAACCCGATT AACGGGCATG CTAAGTTTGT TGCTCGATTC CTCGAAGAAG 1920
 30 AATTGAATAT AACCAATTTT AATCTGATAG AATTATATCA TAATTTGCTT ATTGGCAAAC 1980
 TAGACTCCTA TCTAGATCGT TGGTCAGCAT GTAAAGAGTA TAAGGATCGG CTTCAATTTTG 2040
 35 AACCCATTGA TACAATTTTT CAAGAGCTAT TTAATTTGGA CAACAGAAGT GGATTACTTA 2100
 CCCAGTCGAT TTTCCCTTCT TACAAGTCAA ATATCGAAGA TAACTTACTA AGTTGGGAGC 2160
 AGGTGCTGCC TTCGCTTGAT AAAGAAAATT ATGATACTCT TTCTGGAGAT TTGGATAAAA 2220
 40 TAATGGCTCC GGTACTGGGT CAGCTATTCA AGCTTTATCG TGAGGCGAAT ATGACTATCA 2280
 ACATTTACGA TTTCTACATT GCGTTCAGAG AAACATTACC AAAAGAGGAA ATATTAAATT 2340
 45 TCATAAGAAA AGATCCCTCC AACACCAAAC TCTTAGAACT AGCAGAAACA CCGGACGCAT 2400
 TTGACAAAGT AGCACTAATT TTATTCATGC AAGCAATCTT CGCCTTTGAA AACATGGGTC 2460
 TCATTAAGTT TCAAAGCACC AAGAGTTACG ATCTGGTAGA AAAATGTGTC TGGAGAGGAA 2520
 50 TTTAGATAAA GAATGCACGG ATAAATAAGT AAATAAATAA CCATACATAT ATAGAACCAT 2580
 AGAACCACGT TTTTGTAATG AACAGTCTAC CTGTATCTCA TCATTTTCT GTGTAACTA 2640
 55 TTATTATTAT TATTATCGAA TGGAGGGTAA TATTATGTAT AGGTAAAATA AATAGATAGT 2700
 GCCATGATGC GCGAAGATTG GCAATGGGAA ACTCAAGAAG GCAGCAACAA AAAAATAAA 2759

60 (2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 615 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

65

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

5	Met	Ser	Asp	Leu	Asn	Gln	Ser	Lys	Lys	Met	Asn	Val	Ser	Glu	Phe	Ala	1	5	10	15
	Asp	Ala	Gln	Arg	Ser	His	Tyr	Thr	Val	Tyr	Pro	Ser	Leu	Pro	Gln	Ser	20	25	30	
10	Asn	Lys	Asn	Asp	Lys	His	Ile	Pro	Phe	Val	Lys	Leu	Leu	Ser	Gly	Lys	35	40	45	
	Glu	Ser	Glu	Val	Asn	Val	Glu	Lys	Arg	Trp	Glu	Leu	Tyr	His	Gln	Leu	50	55	60	
15	His	Ser	His	Phe	His	Asp	Gln	Val	Asp	His	Ile	Ile	Asp	Asn	Ile	Glu	65	70	75	80
	Ala	Asp	Leu	Lys	Ala	Glu	Ile	Ser	Asp	Leu	Leu	Tyr	Ser	Glu	Thr	Thr	85	90	95	
20	Gln	Lys	Arg	Arg	Cys	Phe	Asn	Thr	Ile	Phe	Leu	Leu	Gly	Ser	Asp	Ser	100	105	110	
	Thr	Thr	Lys	Ile	Glu	Leu	Lys	Asp	Glu	Ser	Ser	Arg	Tyr	Asn	Val	Leu	115	120	125	
25	Ile	Glu	Leu	Thr	Pro	Lys	Glu	Ser	Pro	Asn	Val	Arg	Met	Met	Leu	Arg	130	135	140	
30	Arg	Ser	Met	Tyr	Lys	Leu	Tyr	Ser	Ala	Ala	Asp	Ala	Glu	Glu	His	Pro	145	150	155	160
	Thr	Ile	Lys	Tyr	Glu	Asp	Ile	Asn	Asp	Glu	Asp	Gly	Asp	Phe	Thr	Glu	165	170	175	
35	Gln	Asn	Asn	Asp	Val	Ser	Tyr	Asp	Leu	Ser	Leu	Val	Glu	Asn	Phe	Lys	180	185	190	
40	Arg	Leu	Phe	Gly	Lys	Asp	Leu	Ala	Met	Val	Phe	Asn	Phe	Lys	Asp	Val	195	200	205	
	Asp	Ser	Ile	Asn	Phe	Asn	Thr	Leu	Asp	Asn	Phe	Ile	Ile	Leu	Leu	Lys	210	215	220	
45	Ser	Ala	Phe	Lys	Tyr	Asp	His	Val	Lys	Ile	Ser	Leu	Ile	Phe	Asn	Ile	225	230	235	240
	Asn	Thr	Asn	Leu	Ser	Asn	Ile	Glu	Lys	Asn	Leu	Arg	Gln	Ser	Thr	Ile	245	250	255	
50	Arg	Leu	Leu	Lys	Arg	Asn	Tyr	His	Lys	Leu	Asp	Val	Ser	Ser	Asn	Lys	260	265	270	
	Gly	Phe	Lys	Tyr	Gly	Asn	Gln	Ile	Phe	Gln	Ser	Phe	Leu	Asp	Thr	Val	275	280	285	
55	Asp	Gly	Lys	Leu	Asn	Leu	Ser	Asp	Arg	Phe	Val	Glu	Phe	Ile	Leu	Ser	290	295	300	
60	Lys	Met	Ala	Asn	Asn	Thr	Asn	His	Asn	Leu	Gln	Leu	Leu	Thr	Lys	Met	305	310	315	320
	Leu	Asp	Tyr	Ser	Leu	Met	Ser	Tyr	Phe	Phe	Gln	Asn	Ala	Phe	Ser	Val	325	330	335	
65	Phe	Ile	Asp	Pro	Val	Asn	Val	Asp	Phe	Leu	Asn	Asp	Asp	Tyr	Leu	Lys	340	345	350	

Ile Leu Ser Arg Cys Pro Thr Phe Met Phe Phe Val Glu Gly Leu Ile
 355 360 365
 5 Lys Gln His Ala Pro Ala Asp Glu Ile Leu Ser Leu Leu Thr Asn Lys
 370 375 380
 Asn Arg Gly Leu Glu Glu Phe Phe Val Glu Phe Leu Val Arg Glu Asn
 385 390 395 400
 10 Pro Ile Asn Gly His Ala Lys Phe Val Ala Arg Phe Leu Glu Glu Glu
 405 410 415
 15 Leu Asn Ile Thr Asn Phe Asn Leu Ile Glu Leu Tyr His Asn Leu Leu
 420 425 430
 Ile Gly Lys Leu Asp Ser Tyr Leu Asp Arg Trp Ser Ala Cys Lys Glu
 435 440 445
 20 Tyr Lys Asp Arg Leu His Phe Glu Pro Ile Asp Thr Ile Phe Gln Glu
 450 455 460
 Leu Phe Thr Leu Asp Asn Arg Ser Gly Leu Leu Thr Gln Ser Ile Phe
 465 470 475 480
 25 Pro Ser Tyr Lys Ser Asn Ile Glu Asp Asn Leu Leu Ser Trp Glu Gln
 485 490 495
 30 Val Leu Pro Ser Leu Asp Lys Glu Asn Tyr Asp Thr Leu Ser Gly Asp
 500 505 510
 Leu Asp Lys Ile Met Ala Pro Val Leu Gly Gln Leu Phe Lys Leu Tyr
 515 520 525
 35 Arg Glu Ala Asn Met Thr Ile Asn Ile Tyr Asp Phe Tyr Ile Ala Phe
 530 535 540
 Arg Glu Thr Leu Pro Lys Glu Glu Ile Leu Asn Phe Ile Arg Lys Asp
 545 550 555 560
 40 Pro Ser Asn Thr Lys Leu Leu Glu Leu Ala Glu Thr Pro Asp Ala Phe
 565 570 575
 Asp Lys Val Ala Leu Ile Leu Phe Met Gln Ala Ile Phe Ala Phe Glu
 580 585 590
 45 Asn Met Gly Leu Ile Lys Phe Gln Ser Thr Lys Ser Tyr Asp Leu Val
 595 600 605
 50 Glu Lys Cys Val Trp Arg Gly
 610 615

(2) INFORMATION FOR SEQ ID NO:7:

55 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2404 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

60 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

65 CTCGAGGCCA CCAAGAAGAG AAAGAGAAGA GCCAGATATT GACTGGAGTG CAGCCAGAGG 60
 TTCCAACCTC CAAAGCTCCT CGGAGCCACC AAGAAGAGAA AGAGAAAAGG AAGAACCAGC 120

TTTGGATTGG GGTGCTGCCA GAGGTGCTCA GTTTGGTAAG CCTCAACAAA CCAAAAATAC 180
 CTACAAGGAT AGGTCTCTAA CTAACAAAAA GACTACTGAT GAGCAACCAA AAATCCAGAA 240
 5 GTCTGTTTAT GATGTTTTAC GTACTGAAGA TGATGATGAA GATGAAGAGG CTGAAAAGCA 300
 AAATGGAGAC GCAAAAGAAA ACAAAGTTGA TCGGGCAGTT GAAAAGCTAC AGGATAAAAC 360
 10 TGCTCAATTG ACTGTTGAAG ATGGTGACAA TTGGGAAGTT GTTGGTAAGA AATAGAGTGT 420
 TGTATGATGA TAAAATGTAC ATTTGTATTT ACTGTTTGCT TTTTTTCTTT CTTGTTTTTTC 480
 TACTCTCCTT TCTACCAGGT ATTCTAACTC TATTATATAA TTAAAAAAA AATAACCATA 540
 15 TATTTTGTAT TAAGTTTCAT ACATGTGTTT AAGTGTATTT TTGGATTTAT CATTTTTTCTA 600
 TGTGAGGTAA GTTTTTGAAT GTCCCATTTT CCTTTCGTTT TTGGAAAGTT CTAAGAAAAA 660
 20 GCATTAACAA TTAATAAAAA AAAAAAATC TAAATAATAC TGATAGAAAT ATCAAATATA 720
 AACTACTAAT ATCGGTAATA TTCAAAGAA GAAGCATGAC TATAAGCGAA GCTCGTCTAT 780
 CACCGCAAGT CAATCTTCTC CCAATAAAGA GGCAC TCAA CGAAGAGGTA GAGGAGACTG 840
 25 CAGCGATTCT AAAAAAGCGT ACTATAGATA ATGAAAAGTG TAAAGACAGC GACCCTGGTT 900
 TTGGTTCCCT TCAAAGAAGG TTACTGCAGC AACTTTATGG CACACTTCCT ACGGACGAAA 960
 30 AGATAATCTT CACATATTTA CAAGATTGTC AACAAGAGAT CGATAGAATC ATTAAACAAT 1020
 CCATTATTCA GAAAGAGAGT CATTGAGTAA TTCTCGTGGG GCCCAGACAA AGTTACAAAA 1080
 CATACTTATT AGACTATGAA CTGTCTTTGT TGCAACAATC TTATAAGAG CAGTTTATAA 1140
 35 CTATCAGGTT GAATGGGTTT ATTCACTCCG AACAAACAGC TATTAACGGT ATAGCAACTC 1200
 AATTGGAACA GCAGTTGCAG AAAATTCATG GCAGTGAAGA AAAAATTGAC GATACTTCAT 1260
 40 TAGAGACTAT TAGCAGTGGT TCTTTGACAG AAGTGTGTTGA GAAAATTCTT TTACTCTTAG 1320
 ATTCGACCAC GAAGACAAGA AATGAAGATA GTGGTGAGGT TGACAGAGAG AGTATAACAA 1380
 AGATAACAGT TGTTTTTATA TTCGATGAAA TTGATACATT TGCTGGGCCT GTGAGGCAAA 1440
 45 CTTTATTATA CAATCTTTTT GACATGGTAG AACATTCTCG GGTACCTGTT TGCATTTTTG 1500
 GCTGCACAAC GAAATTAAAT ATCTTGAAT ATTTAGAAAA GAGGGTAAAG AGTAGATTTT 1560
 50 CTCAAAGAGT GATTTATATG CCGCAAATAC AGAATCTAGA CGATATGGTT GACGCCGTCA 1620
 GAAATTTACT TACAGTTCGC TCTGAAATCT CCCCTGGGT TTCACAATGG AATGAAACGT 1680
 TGGAAAAAGA ACTATCCGAC CCTCGATCGA ATTTGAATAG ACATATTAGG ATGAATTTTCG 1740
 55 AAACCTTTAG GTCATTACCT ACATTGAAAA ATAGCATAAT TCCATTAGTA GCGACATCCA 1800
 AAAATTTTGG TTCACTCTGC ACTGCCATAA AATCGTGTTT TTTTCTTGAC ATATAAATA 1860
 60 AGAACCAACT ATCTAATAAT TTAACAGGAA GGCTCCAATC TTTATCCGAT TTAGAGTTAG 1920
 CCATTTTGAT CTCAGCCGCT AGGGTTGCCT TAAGGGCGAA AGACGGATCT TTTAATTTTA 1980
 ATTTAGCTTA TGCAGAGTAT GAAAAGATGA TTAAAGCTAT CAACTCCAGA ATTCCCACCG 2040
 65 TGGCTCCTAC TACAAATGTG GGAACAGGTC AAAGTACTTT TTCTATCGAC AATACTATCA 2100
 AACTATGGTT GAAAAGGAC GTCAAGAACG TTTGGGAAAA TTTAGTGCAA CTGGATTTTT 2160

TTACCGAGAA ATCAGCCGTT GGT TTGAGAG ATAATGCGAC CGCAGCATT TACGCTAGCA 2220
 ATTATCAATT TCAGGGCACC ATGATCCCGT TTGACTTGAG AAGTTACCAG ATGCAGATCA 2280
 5 TTCTTCAGGA ATTAAGAAGA ATTATCCCCA AATCTAATAT GTACTACTCC TGGACACAAC 2340
 TGTGAATCTT GCGAACAATA TACAGACATT TTATTGGCGG TAGCAACTCT GATATTCCAC 2400
 TGT T 2404

(2) INFORMATION FOR SEQ ID NO:8:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 529 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

25 Met Thr Ile Ser Glu Ala Arg Leu Ser Pro Gln Val Asn Leu Leu Pro
 1 5 10 15
 Ile Lys Arg His Ser Asn Glu Glu Val Glu Glu Thr Ala Ala Ile Leu
 20 25 30
 30 Lys Lys Arg Thr Ile Asp Asn Glu Lys Cys Lys Asp Ser Asp Pro Gly
 35 35 40 45
 Phe Gly Ser Leu Gln Arg Arg Leu Leu Gln Gln Leu Tyr Gly Thr Leu
 50 55 60
 Pro Thr Asp Glu Lys Ile Ile Phe Thr Tyr Leu Gln Asp Cys Gln Gln
 65 70 75 80
 40 Glu Ile Asp Arg Ile Ile Lys Gln Ser Ile Ile Gln Lys Glu Ser His
 85 90 95
 Ser Val Ile Leu Val Gly Pro Arg Gln Ser Tyr Lys Thr Tyr Leu Leu
 100 105 110
 45 Asp Tyr Glu Leu Ser Leu Leu Gln Gln Ser Tyr Lys Glu Gln Phe Ile
 115 120 125
 50 Thr Ile Arg Leu Asn Gly Phe Ile His Ser Glu Gln Thr Ala Ile Asn
 130 135 140
 Gly Ile Ala Thr Gln Leu Glu Gln Gln Leu Gln Lys Ile His Gly Ser
 145 150 155 160
 55 Glu Glu Lys Ile Asp Asp Thr Ser Leu Glu Thr Ile Ser Ser Gly Ser
 165 170 175
 Leu Thr Glu Val Phe Glu Lys Ile Leu Leu Leu Leu Asp Ser Thr Thr
 180 185 190
 60 Lys Thr Arg Asn Glu Asp Ser Gly Glu Val Asp Arg Glu Ser Ile Thr
 195 200 205
 65 Lys Ile Thr Val Val Phe Ile Phe Asp Glu Ile Asp Thr Phe Ala Gly
 210 215 220
 Pro Val Arg Gln Thr Leu Leu Tyr Asn Leu Phe Asp Met Val Glu His
 225 230 235 240

Ser Arg Val Pro Val Cys Ile Phe Gly Cys Thr Thr Lys Leu Asn Ile
 245 250 255
 5 Leu Glu Tyr Leu Glu Lys Arg Val Lys Ser Arg Phe Ser Gln Arg Val
 260 265 270
 Ile Tyr Met Pro Gln Ile Gln Asn Leu Asp Asp Met Val Asp Ala Val
 275 280 285
 10 Arg Asn Leu Leu Thr Val Arg Ser Glu Ile Ser Pro Trp Val Ser Gln
 290 295 300
 Trp Asn Glu Thr Leu Glu Lys Glu Leu Ser Asp Pro Arg Ser Asn Leu
 305 310 315 320
 15 Asn Arg His Ile Arg Met Asn Phe Glu Thr Phe Arg Ser Leu Pro Thr
 325 330 335
 20 Leu Lys Asn Ser Ile Ile Pro Leu Val Ala Thr Ser Lys Asn Phe Gly
 340 345 350
 Ser Leu Cys Thr Ala Ile Lys Ser Cys Ser Phe Leu Asp Ile Tyr Asn
 355 360 365
 25 Lys Asn Gln Leu Ser Asn Asn Leu Thr Gly Arg Leu Gln Ser Leu Ser
 370 375 380
 Asp Leu Glu Leu Ala Ile Leu Ile Ser Ala Ala Arg Val Ala Leu Arg
 385 390 395 400
 30 Ala Lys Asp Gly Ser Phe Asn Phe Asn Leu Ala Tyr Ala Glu Tyr Glu
 405 410 415
 35 Lys Met Ile Lys Ala Ile Asn Ser Arg Ile Pro Thr Val Ala Pro Thr
 420 425 430
 Thr Asn Val Gly Thr Gly Gln Ser Thr Phe Ser Ile Asp Asn Thr Ile
 435 440 445
 40 Lys Leu Trp Leu Lys Lys Asp Val Lys Asn Val Trp Glu Asn Leu Val
 450 455 460
 Gln Leu Asp Phe Phe Thr Glu Lys Ser Ala Val Gly Leu Arg Asp Asn
 465 470 475 480
 45 Ala Thr Ala Ala Phe Tyr Ala Ser Asn Tyr Gln Phe Gln Gly Thr Met
 485 490 495
 50 Ile Pro Phe Asp Leu Arg Ser Tyr Gln Met Gln Ile Ile Leu Gln Glu
 500 505 510
 Leu Arg Arg Ile Ile Pro Lys Ser Asn Met Tyr Tyr Ser Trp Thr Gln
 515 520 525
 55 Leu

(2) INFORMATION FOR SEQ ID NO:9:

- 60 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2306 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 65 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GCTATTTTTT CATGCGTCAG ATGTCACAAA GCCTTTAATC AAGTATTGTT GCAAGAACAC 60
 5 CTGATTCAAA AACTACGTTT TGATATCGAA TCCTATTTAA TTCAAGATTG GAGATGCTCC 120
 AGATGTCATA AAGTGAAACG TGAATATATG AGTGCCCACT GTCCATGTGC CGGCGCGTGG 180
 10 GAAGGAACTC TCCCCAGAGA AAGCATTGTT CAAAAGTTAA ATGTGTTTAA GCAAGTAGCC 240
 AAGTATTACG GTTTTGATAT ATTATTGAGT TGTATTGCTG ATTTGACCAT ATGAGTAAGC 300
 AGTATATAAC GCGAGGTTCA ATGGCCTCTT TACCATGAAA AAAAAAAAAA AAAAAAAAAA 360
 15 AAGGTAAGGA AAAAGAGTAT TTTCAATTCG TTTCTGAACA TATAAATATA AATAACCGAA 420
 AAATTAGCCC TTGAACATAA TTAACACTCT TCTTTGATAT TTAAATCACA AGTACTTTTC 480
 20 TTTTATTTTC TTCTTAATAC TTTTGGAAT AAAATGAATG TGACCACTCC GGAAGTTGCT 540
 TTTAGGGAAT ATCAAACCAA CTGTCTCGCA TCGTATATTT CTGCTGATCC AGACATAACT 600
 CCTTCAAATT TAATCTTGCA AGGTTATAGT GGAACAGGAA AAACCTACAC TTTGAAGAAG 660
 25 TATTTTAATG CGAATCCAAA TTTGCATGCA GTATGGCTGG AACCTGTTGA GTTGGTTTCT 720
 TGGAAGCCCT TACTGCAGGC GATAGCACGT ACTGTACAAT ATAAATTGAA AACCCCTATAT 780
 CCAAACATTC CCACCACAGA TTACGATCCT TTACAGGTTG AAGAGCCATT TCTTTTGGTA 840
 30 AAGACGTTGC ACAATATTTT TGTCCAATAT GAATCTTTGC AAGAAAAGAC TTGCTTGTTT 900
 TTGATATTGG ATGGTTTCGA TAGTTTACAA GATTAGACG CCGCACTGTT TAACAAATAT 960
 35 ATCAAATAA ATGAATTACT TCCAAAAGAT TCTAAAATTA ATATAAAATT CATTTACACG 1020
 ATGTTAGAGA CATCATTTTT GCAAAGATAT TCTACACATT GCATTCCAAC TGTTATGTTT 1080
 CCGAGGTATA ATGTGGACGA AGTTTCTACT ATATTAGTGA TGTCTAGATG TGGCGAACTC 1140
 40 ATGGAAGATT CTTGTCTACG TAAGCGTATC ATTGAAGAGC AGATAACGGA CTGTACAGAC 1200
 GATCAATTTT AAAATGTAGC TGCGAACTTC ATTCACTTAA TTGTGCAGGC TTTTCATTCT 1260
 45 TATACTGGAA ACGACATATT CGCATTGAAT GACTTGATAG ACTTCAAATG GCCCAAGTAT 1320
 GTATCTCGCA TTACTAAGGA AAACATATTT GAACCACTGG CTCTTTACAA AAGTGCCATC 1380
 50 AAATATTTT TAAGCACAGA TGATAATTTA AGTGAAAATG GACAAGGTGA AAGCGCGATA 1440
 ACCACAAATC GTGATGACCT TGAGAACAGT CAACTTACG ACTTATCAAT AATTTCAAG 1500
 TATCTGCTCA TAGCCTCATA TATTTGTTCA TATCTGGAAC CTAGATACGA TGCGAGTATT 1560
 55 TTCTCTAGGA AAACACGTAT CATAAAGGT AGAGCTGCTT ATGGACGAAG AAAGAAGAAA 1620
 GAAGTTAACC CTAGATATTT ACAGCCTTCT TTATTGCTA TTGAAAGACT TTTGGCTATT 1680
 60 TTCCAAGCTA TATCCCTAT TCAAGGTAAG GCGGAGAGTG GTTCCCTATC TGCACTTCGT 1740
 GAGGAATCCT TAATGAAAGC GAATATCGAG GTTTTCAAAA ATTTATCCGA ATTGCATACA 1800
 TTGAAATTAA TAGCTACAAC CATGAACAAG AATATCGACT ATTTGAGTCC TAAAGTCAGG 1860
 65 TGGAAAGTAA ACGTTCCCTG GGAAATTATT AAAGAAATAT CAGAATCTGT TCATTTCAT 1920
 ATCAGCGATT ACTTCAGCGA TATTCACGAA TGATTATCTC CCTGGAAGGT ATCCAGAGGG 1980

CAGGATACGT TCGAAACAAC AACTACGTTA TATAAATATT TATACATAGT GGGATAGAAT 2040
 GAACAATTAT CAAGTAAACC TTGTATTTTT TGTTCCCACG CTCTACGCTC TGTTTCTTGG 2100
 5 ATATGGTAAT CAAAGATTAA TACGTATAAC CGTTATTAAT TCAGTCCACT AGAACTATT 2160
 AAAAGCGCCC TACTGTATGG AAAAACAATG AATGAGGAGA CTGAACGGCG CAAAATTGTT 2220
 AGTTTAGTTG CTCTTTTGG CGGCCGGCGA TAATGTTCTT CACTTGGTAT TCTTACCAGG 2280
 10 ATTGAGCCTG ATTTTGTTTT GTCTTA 2306

(2) INFORMATION FOR SEQ ID NO:10:

15

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 479 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 20 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Asn Val Thr Thr Pro Glu Val Ala Phe Arg Glu Tyr Gln Thr Asn
 1 5 10 15
 Cys Leu Ala Ser Tyr Ile Ser Ala Asp Pro Asp Ile Thr Pro Ser Asn
 20 25 30
 Leu Ile Leu Gln Gly Tyr Ser Gly Thr Gly Lys Thr Tyr Thr Leu Lys
 35 40 45
 Lys Tyr Phe Asn Ala Asn Pro Asn Leu His Ala Val Trp Leu Glu Pro
 50 55 60
 Val Glu Leu Val Ser Trp Lys Pro Leu Leu Gln Ala Ile Ala Arg Thr
 65 70 75 80
 Val Gln Tyr Lys Leu Lys Thr Leu Tyr Pro Asn Ile Pro Thr Thr Asp
 85 90 95
 Tyr Asp Pro Leu Gln Val Glu Glu Pro Phe Leu Leu Val Lys Thr Leu
 100 105 110
 His Asn Ile Phe Val Gln Tyr Glu Ser Leu Gln Glu Lys Thr Cys Leu
 115 120 125
 Phe Leu Ile Leu Asp Gly Phe Asp Ser Leu Gln Asp Leu Asp Ala Ala
 130 135 140
 Leu Phe Asn Lys Tyr Ile Lys Leu Asn Glu Leu Leu Pro Lys Asp Ser
 145 150 155 160
 Lys Ile Asn Ile Lys Phe Ile Tyr Thr Met Leu Glu Thr Ser Phe Leu
 165 170 175
 Gln Arg Tyr Ser Thr His Cys Ile Pro Thr Val Met Phe Pro Arg Tyr
 180 185 190
 Asn Val Asp Glu Val Ser Thr Ile Leu Val Met Ser Arg Cys Gly Glu
 195 200 205
 65 Leu Met Glu Asp Ser Cys Leu Arg Lys Arg Ile Ile Glu Glu Gln Ile
 210 215 220

Thr Asp Cys Thr Asp Asp Gln Phe Gln Asn Val Ala Ala Asn Phe Ile
 225 230 235 240
 5 His Leu Ile Val Gln Ala Phe His Ser Tyr Thr Gly Asn Asp Ile Phe
 245 250 255
 Ala Leu Asn Asp Leu Ile Asp Phe Lys Trp Pro Lys Tyr Val Ser Arg
 260 265 270
 10 Ile Thr Lys Glu Asn Ile Phe Glu Pro Leu Ala Leu Tyr Lys Ser Ala
 275 280 285
 Ile Lys Leu Phe Leu Ser Thr Asp Asp Asn Leu Ser Glu Asn Gly Gln
 290 295 300
 15 Gly Glu Ser Ala Ile Thr Thr Asn Arg Asp Asp Leu Glu Asn Ser Gln
 305 310 315 320
 Thr Tyr Asp Leu Ser Ile Ile Ser Lys Tyr Leu Leu Ile Ala Ser Tyr
 325 330 335
 20 Ile Cys Ser Tyr Leu Glu Pro Arg Tyr Asp Ala Ser Ile Phe Ser Arg
 340 345 350
 Lys Thr Arg Ile Ile Gln Gly Arg Ala Ala Tyr Gly Arg Arg Lys Lys
 355 360 365
 Lys Glu Val Asn Pro Arg Tyr Leu Gln Pro Ser Leu Phe Ala Ile Glu
 370 375 380
 30 Arg Leu Leu Ala Ile Phe Gln Ala Ile Phe Pro Ile Gln Gly Lys Ala
 385 390 395 400
 Glu Ser Gly Ser Leu Ser Ala Leu Arg Glu Glu Ser Leu Met Lys Ala
 405 410 415
 35 Asn Ile Glu Val Phe Gln Asn Leu Ser Glu Leu His Thr Leu Lys Leu
 420 425 430
 Ile Ala Thr Thr Met Asn Lys Asn Ile Asp Tyr Leu Ser Pro Lys Val
 435 440 445
 Arg Trp Lys Val Asn Val Pro Trp Glu Ile Ile Lys Glu Ile Ser Glu
 450 455 460
 45 Ser Val His Phe Asn Ile Ser Asp Tyr Phe Ser Asp Ile His Glu
 465 470 475

50 (2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1975 base pairs
 (B) TYPE: nucleic acid
 55 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

60 (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 443..1747

65 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CGTGTGCTCT TCTATAGTAA TTTGACATTC TCTAAACGCA GAGACCTCTT ATAAAGATTC 60
 AACAAATAAG GAATGTTACC TATGCTAGTC GCAACTCTCT CGTAAGTTGA GGGTTGCTAA 120

	CAGAAAAACG	ATGAGAAGAA	ACTTTTGAAA	AATATTGTGT	GAAAGCAGCA	CGAAACAGAG	180										
	TATGAAAAAA	GAATGCGGGC	GTCCGTAAAG	AGCTAGAATC	GCAAGTGTCC	AGAATATGCA	240										
5	AGGCTTTTCGA	ATACACTCCT	CACGCTTCTC	TTCAGCAAAA	ATCAACTCTT	TGTGATAAAA	300										
	CTGTGTATTT	CTTTGTTCTT	TGCCGTTGTT	TACGTTAGTA	AGAAATCGGC	ATTGAAAAAA	360										
10	AAAATCTCAC	ACTAAAATTG	CAGAAAAAAG	TGTACAATAT	CAGTAAATAA	AATTGGCCAA	420										
	AACAATACCA	TTAAAACCAG	TC	ATG	TCC	ATG	CAA	CAA	GTC	CAA	CAT	TGT	GTC	472			
				Met	Ser	Met	Gln	Gln	Val	Gln	His	Cys	Val	10			
				1					5								
15	GCA	GAA	GTA	CTT	CGA	CTA	GAT	CCA	CAA	GAA	AAA	CCG	GAC	TGG	TCG	AGC	520
	Ala	Glu	Val	Leu	Arg	Leu	Asp	Pro	Gln	Glu	Lys	Pro	Asp	Trp	Ser	Ser	
					15					20					25		
20	GGA	TAT	TTG	AAG	AAG	TTG	ACT	AAT	GCG	ACA	TCG	ATT	TTA	TAT	AAT	ACT	568
	Gly	Tyr	Leu	Lys	Lys	Leu	Thr	Asn	Ala	Thr	Ser	Ile	Leu	Tyr	Asn	Thr	
				30					35					40			
25	TCA	CTG	AAC	AAG	GTA	ATG	CTG	AAA	CAA	GAT	GAA	GAG	GTT	GCT	AGA	TGT	616
	Ser	Leu	Asn	Lys	Val	Met	Leu	Lys	Gln	Asp	Glu	Glu	Val	Ala	Arg	Cys	
			45					50					55				
30	CAC	ATA	TGT	GCA	TAC	ATA	GCG	TCA	CAG	AAA	ATG	AAT	GAA	AAA	CAC	ATG	664
	His	Ile	Cys	Ala	Tyr	Ile	Ala	Ser	Gln	Lys	Met	Asn	Glu	Lys	His	Met	
		60					65					70					
35	CCT	GAC	CTT	TGC	TAT	TAT	ATA	GAC	AGT	ATT	CCC	TTG	GAG	CCG	AAA	AAA	712
	Pro	Asp	Leu	Cys	Tyr	Tyr	Ile	Asp	Ser	Ile	Pro	Leu	Glu	Pro	Lys	Lys	
	75					80					85					90	
40	GCC	AAG	CAT	TTA	ATG	AAC	CTT	TTC	AGA	CAA	AGT	TTA	TCT	AAT	TCT	TCA	760
	Ala	Lys	His	Leu	Met	Asn	Leu	Phe	Arg	Gln	Ser	Leu	Ser	Asn	Ser	Ser	
					95					100					105		
45	CCT	ATG	AAA	CAA	TTT	GCT	TGG	ACA	CCG	AGC	CCC	AAA	AAG	AAC	AAA	CGC	808
	Pro	Met	Lys	Gln	Phe	Ala	Trp	Thr	Pro	Ser	Pro	Lys	Lys	Asn	Lys	Arg	
				110					115					120			
50	AGT	CCA	GTA	AAG	AAC	GGT	GGG	AGG	TTT	ACT	TCT	TCT	GAT	CCG	AAA	GAG	856
	Ser	Pro	Val	Lys	Asn	Gly	Gly	Arg	Phe	Thr	Ser	Ser	Asp	Pro	Lys	Glu	
			125					130					135				
55	TTG	AGG	AAT	CAA	CTG	TTT	GGT	ACA	CCA	ACT	AAA	GTT	AGG	AAA	AGC	CAA	904
	Leu	Arg	Asn	Gln	Leu	Phe	Gly	Thr	Pro	Thr	Lys	Val	Arg	Lys	Ser	Gln	
		140					145					150					
60	AAT	AAT	GAT	TCG	TTC	GTA	ATA	CCA	GAA	CTA	CCC	CCC	ATG	CAA	ACC	AAT	952
	Asn	Asn	Asp	Ser	Phe	Val	Ile	Pro	Glu	Leu	Pro	Pro	Met	Gln	Thr	Asn	
	155					160					165					170	
65	GAA	TCG	CCG	TCT	ATT	ACT	AGG	AGA	AAG	TTA	GCA	TTT	GAA	GAG	GAT	GAG	1000
	Glu	Ser	Pro	Ser	Ile	Thr	Arg	Arg	Lys	Leu	Ala	Phe	Glu	Glu	Asp	Glu	
					175					180					185		
70	GAT	GAG	GAT	GAA	GAG	GAA	CCA	GGA	AAC	GAC	GGT	TTG	TCT	TTA	AAA	AGC	1048
	Asp	Glu	Asp	Glu	Glu	Glu	Pro	Gly	Asn	Asp	Gly	Leu	Ser	Leu	Lys	Ser	
				190					195					200			
75	CAT	AGT	AAT	AAG	AGC	ATT	ACT	GGA	ACC	AGA	AAT	GTA	GAT	TCT	GAT	GAG	1096
	His	Ser	Asn	Lys	Ser	Ile	Thr	Gly	Thr	Arg	Asn	Val	Asp	Ser	Asp	Glu	
			205					210					215				
80	TAT	GAA	AAC	CAT	GAA	AGT	GAC	CCT	ACA	AGT	GAG	GAA	GAG	CCA	TTA	GGT	1144

	Tyr	Glu	Asn	His	Glu	Ser	Asp	Pro	Thr	Ser	Glu	Glu	Glu	Pro	Leu	Gly	
	220						225					230					
5	GTG	CAA	GAA	AGC	AGA	AGC	GGG	AGA	ACG	AAA	CAA	AAT	AAG	GCA	GTT	GGA	1192
	Val	Gln	Glu	Ser	Arg	Ser	Gly	Arg	Thr	Lys	Gln	Asn	Lys	Ala	Val	Gly	
	235					240					245				250		
10	AAA	CCG	CAA	TCA	GAA	TTG	AAG	ACG	GCA	AAA	GCC	CTG	AGG	AAA	AGG	GGC	1240
	Lys	Pro	Gln	Ser	Glu	Leu	Lys	Thr	Ala	Lys	Ala	Leu	Arg	Lys	Arg	Gly	
					255					260					265		
15	AGA	ATA	CCA	AAT	TCT	TTG	TTA	GTA	AAG	AAG	TAT	TGC	AAA	ATG	ACT	ACT	1288
	Arg	Ile	Pro	Asn	Ser	Leu	Leu	Val	Lys	Lys	Tyr	Cys	Lys	Met	Thr	Thr	
					270				275					280			
20	GAA	GAA	ATA	ATA	CGG	CTT	TGC	AAC	GAT	TTT	GAA	TTA	CCA	AGA	GAA	GTA	1336
	Glu	Glu	Ile	Ile	Arg	Leu	Cys	Asn	Asp	Phe	Glu	Leu	Pro	Arg	Glu	Val	
					285			290					295				
25	GCA	TAT	AAA	ATT	GTG	GAT	GAG	TAC	AAC	ATA	AAC	GCG	TCA	AGA	TTG	GTT	1384
	Ala	Tyr	Lys	Ile	Val	Asp	Glu	Tyr	Asn	Ile	Asn	Ala	Ser	Arg	Leu	Val	
		300				305						310					
30	TGC	CCA	TGG	CAA	TTA	GTG	TGT	GGG	TTA	GTA	TTA	AAT	TGT	ACA	TTC	ATT	1432
	Cys	Pro	Trp	Gln	Leu	Val	Cys	Gly	Leu	Val	Leu	Asn	Cys	Thr	Phe	Ile	
		315				320					325				330		
35	GTA	TTT	AAT	GAA	AGA	AGA	CGC	AAG	GAT	CCA	AGA	ATT	GAC	CAT	TTT	ATA	1480
	Val	Phe	Asn	Glu	Arg	Arg	Arg	Lys	Asp	Pro	Arg	Ile	Asp	His	Phe	Ile	
					335					340					345		
40	GTC	AGT	AAG	ATG	TGC	AGC	TTG	ATG	TTG	ACG	TCA	AAA	GTG	GAT	GAT	GTT	1528
	Val	Ser	Lys	Met	Cys	Ser	Leu	Met	Leu	Thr	Ser	Lys	Val	Asp	Asp	Val	
					350				355					360			
45	ATT	GAA	TGT	GTA	AAA	TTA	GTG	AAG	GAA	TTA	ATT	ATC	GGT	GAA	AAA	TGG	1576
	Ile	Glu	Cys	Val	Lys	Leu	Val	Lys	Glu	Leu	Ile	Ile	Gly	Glu	Lys	Trp	
					365			370					375				
50	TTC	AGA	GAT	TTG	CAA	ATT	AGG	TAT	GAT	GAT	TTT	GAT	GGC	ATC	AGA	TAC	1624
	Phe	Arg	Asp	Leu	Gln	Ile	Arg	Tyr	Asp	Asp	Phe	Asp	Gly	Ile	Arg	Tyr	
							385					390					
55	GAT	GAA	ATT	ATA	TTT	AGG	AAA	CTG	GGA	TCG	ATG	TTA	CAA	ACC	ACC	AAT	1672
	Asp	Glu	Ile	Ile	Phe	Arg	Lys	Leu	Gly	Ser	Met	Leu	Gln	Thr	Thr	Asn	
						400					405					410	
60	ATT	TTG	GTC	ACA	GAC	GAC	CAG	TAC	AAT	ATT	TGG	AAG	AAA	AGA	ATT	GAA	1720
	Ile	Leu	Val	Thr	Asp	Asp	Gln	Tyr	Asn	Ile	Trp	Lys	Lys	Arg	Ile	Glu	
					415					420					425		
65	ATG	GAT	TTG	GCA	TTA	ACA	GAA	CCT	TTA	TAACATATCC	AGTATTA	ACT					1767
	Met	Asp	Leu	Ala	Leu	Thr	Glu	Pro	Leu								
					430				435								
70	AAAAGTATAT	ATTGACCAA	TACCTGACAT	ATCTTCTAAA	GCATGCCTTT	AGCCCTATAA											1827
75	CGAGCTAATG	TTAGCTCCAT	CTTTGCACTT	ATGATTGGAT	CAGCCCTCAA	ACGCTTTTGT											1887
80	ATCTTTGCAG	CTTCCGCGAA	GGTAGTAGCT	TGAAGTTTTT	CATCCATAGT	TCTTGCTAAA											1947
85	ATTGCAGAAT	CTTCAAACAA	TTCTATGG														1975

65 (2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 435 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Ser Met Gln Gln Val Gln His Cys Val Ala Glu Val Leu Arg Leu
1 5 10 15
Asp Pro Gln Glu Lys Pro Asp Trp Ser Ser Gly Tyr Leu Lys Lys Leu
20 25 30
Thr Asn Ala Thr Ser Ile Leu Tyr Asn Thr Ser Leu Asn Lys Val Met
35 40 45
Leu Lys Gln Asp Glu Glu Val Ala Arg Cys His Ile Cys Ala Tyr Ile
50 55 60
Ala Ser Gln Lys Met Asn Glu Lys His Met Pro Asp Leu Cys Tyr Tyr
65 70 75 80
Ile Asp Ser Ile Pro Leu Glu Pro Lys Lys Ala Lys His Leu Met Asn
85 90 95
Leu Phe Arg Gln Ser Leu Ser Asn Ser Ser Pro Met Lys Gln Phe Ala
100 105 110
Trp Thr Pro Ser Pro Lys Lys Asn Lys Arg Ser Pro Val Lys Asn Gly
115 120 125
Gly Arg Phe Thr Ser Ser Asp Pro Lys Glu Leu Arg Asn Gln Leu Phe
130 135 140
Gly Thr Pro Thr Lys Val Arg Lys Ser Gln Asn Asn Asp Ser Phe Val
145 150 155 160
Ile Pro Glu Leu Pro Pro Met Gln Thr Asn Glu Ser Pro Ser Ile Thr
165 170 175
Arg Arg Lys Leu Ala Phe Glu Glu Asp Glu Asp Glu Asp Glu Glu Glu
180 185 190
Pro Gly Asn Asp Gly Leu Ser Leu Lys Ser His Ser Asn Lys Ser Ile
195 200 205
Thr Gly Thr Arg Asn Val Asp Ser Asp Glu Tyr Glu Asn His Glu Ser
210 215 220
Asp Pro Thr Ser Glu Glu Glu Pro Leu Gly Val Gln Glu Ser Arg Ser
225 230 235 240
Gly Arg Thr Lys Gln Asn Lys Ala Val Gly Lys Pro Gln Ser Glu Leu
245 250 255
Lys Thr Ala Lys Ala Leu Arg Lys Arg Gly Arg Ile Pro Asn Ser Leu
260 265 270
Leu Val Lys Lys Tyr Cys Lys Met Thr Thr Glu Glu Ile Ile Arg Leu
275 280 285
Cys Asn Asp Phe Glu Leu Pro Arg Glu Val Ala Tyr Lys Ile Val Asp
290 295 300
Glu Tyr Asn Ile Asn Ala Ser Arg Leu Val Cys Pro Trp Gln Leu Val
305 310 315 320

Cys Gly Leu Val Leu Asn Cys Thr Phe Ile Val Phe Asn Glu Arg Arg
 325 330 335
 5 Arg Lys Asp Pro Arg Ile Asp His Phe Ile Val Ser Lys Met Cys Ser
 340 345 350
 Leu Met Leu Thr Ser Lys Val Asp Asp Val Ile Glu Cys Val Lys Leu
 355 360 365
 10 Val Lys Glu Leu Ile Ile Gly Glu Lys Trp Phe Arg Asp Leu Gln Ile
 370 375 380
 Arg Tyr Asp Asp Phe Asp Gly Ile Arg Tyr Asp Glu Ile Ile Phe Arg
 385 390 395 400
 15 Lys Leu Gly Ser Met Leu Gln Thr Thr Asn Ile Leu Val Thr Asp Asp
 405 410 415
 20 Gln Tyr Asn Ile Trp Lys Lys Arg Ile Glu Met Asp Leu Ala Leu Thr
 420 425 430
 Glu Pro Leu
 435

WHAT IS CLAIMED IS:

1. A composition comprising an isolated nucleic acid encoding a biologically active unique portion of an ORC polypeptide.
5
2. A composition according to claim 1, wherein said ORC gene is ORC1.
3. A composition according to claim 1, wherein said ORC gene is
10 ORC2.
4. A composition according to claim 1, wherein said ORC gene is ORC3.
- 15 5. A composition according to claim 1, wherein said ORC gene is ORC4.
6. A composition according to claim 1, wherein said ORC gene is ORC5.
20
7. A composition according to claim 1, wherein said ORC gene is ORC6.
8. A composition comprising a recombinant, biologically active unique
25 portion of an ORC protein.
9. A method of identifying an ORC selective agent, said method comprising the steps of:
contacting an agent with a composition according to claim 8;
30 measuring in at least qualitative terms the binding affinity of said agent for said composition.

10. A method for identifying a gene encoding a protein which directly or indirectly associates with a selected DNA sequence, said method comprising the steps of:

transforming an expression library of hybrid proteins into a reporter strain,
5 wherein said library comprises protein-coding sequences fused to a constitutively expressed transcription activation domain and said reporter strain comprises a reporter gene with at least one copy of a selected DNA sequence in its promoter region;

detecting the transcription or translation product of said reporter gene in a
10 clone of said reporter strain;

recovering said clone;

whereby said clone comprises a gene encoding a protein which directly or indirectly associates with said selected DNA sequence.

INTERNATIONAL SEARCH REPORT

I. national application No.

PCT/US94/14563

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C07H 17/00; C07K 2/00, 14/00; C12N 15/00; C12Q 1/68

US CL : 435/6, 320; 530/350; 536/23.1, 23.4, 23.74, 24.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 320; 530/350; 536/23.1, 23.4, 23.74, 24.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

GENBANK, EMBL

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, DIALOG

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Cell, Volume 65, issued 31 May 1991, Irie et al., "SGV1 Encodes a CDC28/cdc2-Related Kinase Required for a G Alpha Subunit-Mediated Adaptive Response to Phereomone in <i>S. cerevisiae</i> ", pages 785-795, see entire document.	1, 5
X	EMBO Journal, Volume 12, Number 10, issued 1993, Coppolecchia et al., "A New Yeast Translation Initiation Factor Suppresses a Mutation in the eIF-4A RNA Helicase", pages 4005-4011, see entire document.	1, 5
X	EMBL Gene Sequence Listing, Accession Number M60416 M36724, issued 19 May 1992, Clark et al.	1, 6

☒ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:	*T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	*Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A*	document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means		
P document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

28 MARCH 1995

Date of mailing of the international search report

10 APR 1995

 Name and mailing address of the ISA/US
 Commissioner of Patents and Trademarks
 Box PCT
 Washington, D.C. 20231

Facsimile No. (703) 305-3230

 Authorized officer *Douglas Gurian-Sherman*
 DOUGLAS GURIAN-SHERMAN

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/14563

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Nature, Volume 366, issued 4 November 1993, Micklem et al., "Yeast Origin Recognition Complex is Involved in DNA Replication and Transcriptional Silencing", pages 87-89, see entire document.	1, 3
Y	Nature, Volume 357, issued 14 May 1992, Bell et al., "ATP-Dependent Recognition of Eukaryotic Origins of DNA Replication by a Multiprotein Complex", pages 128-134, see entire document.	1-10
Y	Nature, Volume 357, issued 14 May 1992, Diffley et al., "Protein-DNA Interactions at a Yeast Replication Origin", pages 169-172, see entire document.	1-10
Y	PCT WO 92/13091 (FOULKES ET AL) 06 August 1992, pages 13-18.	1-10
Y	Proc. Natl. Acad. Sci., USA, Volume 86, issued March 1989, Gould et al., "Use of the DNA Polymerase Chain Reaction for Homology Probing: Isolation of Partial cDNA or Genomic Clones Encoding the Iron-Sulfur Protein of Succinate dehydrogenase from several species", pages 1934-1938, see entire document.	1-10
Y	Biochemistry, Volume 29, issued 1990, Heiger-Bernays et al., "Effects of the Antitumor Drug cis-Diamminedichloroplatinum(II) and Related Platinum Complexes on Eukaryotic DNA Replication", pages 8461-8466, see entire document.	1-10
Y	Cytometry, Volume 12, issued 1991, Hoffman et al., "A New Class of Reversible Cell Cycle Inhibitors", pages 26-32, see entire document.	1-10
Y	J. Molecular Biology, Volume 183, issued 1985, Lathe, "Synthetic Oligonucleotide Probes Deduced From Amino Acid Sequence Data: Theoretical and Practical Considerations", pages 1-12, see entire document.	1-10
Y	J. Biological Chemistry, Volume 262, Number 21, issued 25 July 1987, Matsudaira, "Sequence from Picomole Quantities of Proteins Electrophoretically Blotted onto Polyvinylidene Difluoride Membranes", pages 10035-10038, see entire document.	1-10

